

P. CAZES

Estimation de la statistique de multiplication du premier étage d'un photomultiplicateur à dynodes

Les cahiers de l'analyse des données, tome 3, n° 4 (1978),
p. 393-417

http://www.numdam.org/item?id=CAD_1978__3_4_393_0

© Les cahiers de l'analyse des données, Dunod, 1978, 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/>

ESTIMATION DE LA STATISTIQUE DE MULTIPLICATION DU PREMIER ÉTAGE D'UN PHOTOMULTIPLICATEUR A DYNODES [PHOTOMULTIPLICATEUR]

par P. Cazes (1)

Introduction

Le problème que l'on essaie de résoudre ici, problème qui nous a été soumis par des physiciens, rentre dans un cadre général : ayant l'histogramme d'une variable Y dont la loi $y(x)$ est donnée par :

$$y(x) = \int g(x') k(x', x) dx'$$

$k(x', x)$ étant connu, estimer la fonction $g(x')$.

Après le rappel du modèle physique (§ 1), qui se met sous la forme précédente (§ 2.1), nous indiquons comment après discrétisation à l'aide d'un découpage relativement fin de l'intervalle de variation de $g(x')$, et obtention d'un système d'équations $y_j = \int g(x'_i) \Delta x'_i k(x'_i, x_j)$ (où y_j ($1 \leq j \leq m$) désigne la fréquence (relative) de la $j^{\text{ème}}$ classe de l'histogramme de Y , $k(x'_i, x_j)$ la masse de $k(x', x)$ correspondante, et x'_i le centre du $i^{\text{ème}}$ intervalle de largeur $\Delta x'_i$ associé au découpage adopté), l'on peut résoudre le problème précédent en se plaçant dans l'espace des premiers facteurs de l'analyse des correspondances du tableau k_{IJ} des $k(x'_i, x_j)$ (cf § 2.2 et 2.3).

Au § 3 nous donnons les résultats obtenus sur l'exemple physique, avec les données en notre possession et qui correspondent à dix histogrammes (spectres à électron unique d'un photomultiplicateur) enregistrés dans des conditions identiques, tandis qu'au § 4 on teste la méthode de régression utilisée en partant de lois $g(x')$ supposées connues et en reconstituant l'histogramme associé de Y , histogramme perturbé ou non par une erreur normale.

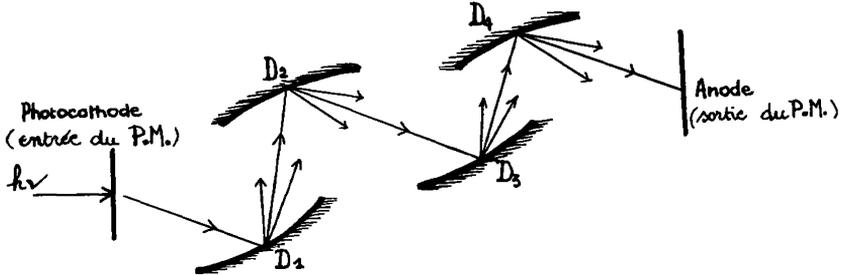
1 Le problème physique

Le spectre d'amplitude des impulsions de sortie d'un photomultiplicateur (P.M.) fonctionnant dans les conditions du spectre à photoélectron unique reflète les caractéristiques de ce P.M. et en particulier de la chaîne de multiplication. Ce spectre qui fournit la probabilité $P(n)$ pour qu'un photoélectron (i.e. un électron libéré par un photon tombant sur la photocathode d'entrée du P.M.) donne naissance à n électrons sur l'anode de sortie du P.M. est encore appelé spectre à électron unique ou s.e.u.. Il peut être mesuré à l'aide d'un analyseur

(1) Maître-assistant, laboratoire de statistique, Université P. & M. Curie
Pour rédiger cet article, nous avons bénéficié de plusieurs entretiens avec
J.P. Ballini et P.Y. Turpin.

multicanaux qui permet de compter le nombre N_n (*) de fois que durant la durée de l'expérience, on a recueilli n électrons sur l'anode de sortie du P.M. ; deux paquets d'électrons associés à deux photoélectrons incidents successifs étant, de par les conditions de l'expérience, assez espacés pour ne pas empiéter, on a bien ainsi le nombre de photoélectrons qui durant la durée de l'expérience ont donné lieu à n électrons.

La chaîne de multiplication est constituée par une série de dynodes, petits "miroirs" métalliques (sous tension) sur lesquels les électrons accélérés par cette tension apportent une énergie suffisante pour extraire de la dynode une gerbe d'électrons secondaires (on dit que les électrons se multiplient).



Principe de la multiplication dans un photomultiplicateur a dynodes

La forme générale du s.e.u. du P.M. est fonction de toute la chaîne de multiplication, mais dépend essentiellement de la statistique de multiplication du premier étage ; c'est la raison pour laquelle on choisit une première dynode de gain supérieur à celui des dynodes suivantes, de façon à avoir une distribution aussi concentrée que possible autour de sa moyenne (i.e. une distribution dont la variance relative, rapport de la variance au carré de la moyenne, soit aussi faible que possible).

Si l'on suppose que la statistique de multiplication à chaque étage est poissonnienne (i.e. que le nombre d'électrons émis à partir d'un électron tombant sur la $i^{\text{ème}}$ dynode et collectée sur la dynode suivante suit une loi de Poisson de paramètre δ_i gain moyen de l'étage), on obtient un s.e.u. théorique moins dispersé (i.e. de variance plus faible) que le s.e.u. expérimental. Aussi d'autres distributions (du moins en ce qui concerne le premier étage dont l'influence est comme on l'a dit ci-dessus, prépondérante pour déterminer la forme du s.e.u.) que la loi de Poisson, ont été proposées pour la statistique de multiplication d'un étage ; ces distributions reviennent à considérer que l'on a une distribution de Poisson dont le paramètre δ' est aléatoire et possède une certaine dispersion de loi $g(\delta')$.

D'un point de vue physique, on peut considérer que la fonction $g(\delta')$ traduit l'inhomogénéité de la dynode D considérée, ainsi que la variabilité de la proportion des électrons issus de D et collectés par la dynode suivante D' (en effet certains électrons issus de D peuvent ne pas être captés par D' et sont donc perdus), proportion qui est encore appelée facteur de collection η (ou $\eta_{DD'}$) entre D et D' ; en effet, suivant le point

(*) En fait, le processus est un peu plus complexe, l'analyseur multicanaux permettant d'enregistrer non pas N_n , mais le nombre de fois qu'à la sortie du P.M., on a une impulsion dont l'amplitude correspond au $n^{\text{ème}}$ canal de cet analyseur (cf § 3.1).

M de D sur lequel tombe un électron, on aura un gain moyen $m(M)$ et un facteur de collection $\eta(M)$ fonctions du point M, et donc un gain de l'étage $\delta(M) = \eta(M) m(M)$ non uniforme. On conçoit donc qu'on puisse avoir un gain variant suivant une loi $g(\delta')$ fonction des conditions expérimentales. La probabilité $f(n)$ qu'à un étage donné, un électron donne naissance à n électrons s'écrit alors :

$$f(n) = \int_0^{\infty} (g(\delta') \exp(-\delta') \delta'^n / n!) d\delta' \quad (1)$$

Si $g(\delta')$ est la mesure de Dirac centrée en δ , gain moyen de l'étage, on retrouve la distribution de Poisson.

Soit μ et σ la moyenne de l'écart-type de la loi $g(\delta')$. La moyenne de $f(n)$ étant aussi égale à μ , on doit avoir $\mu = \delta$, si δ désigne le gain moyen de l'étage étudié. La variance de la loi $f(n)$, égale à $\mu + \sigma^2$, i.e. à $\delta + \sigma^2$ est bien comme prévu plus forte que la variance de la loi de Poisson de paramètre δ .

Si l'on suppose que $a\delta'$ suit une loi gamma d'ordre t

$(g(\delta') = a^t \delta'^{t-1} \exp(-a\delta') / \Gamma(t)$; $\mu = t/a = \delta$; $\sigma^2 = t/a^2 = \delta/a$), on obtient pour $f(n)$ la distribution binomiale négative :

$$f(n) = a^t \Gamma(n+t) / ((1+a)^{n+t} n! \Gamma(t)) \quad (2)$$

Si t tend vers l'infini, compte-tenu du fait que le rapport t/a reste constant et égal à δ , on retrouve la loi de Poisson de paramètre δ . Notons que les paramètres a et t (ou δ et a) sont en général estimés par la méthode des moments (cf [1]).

La loi précédente, qui a été préconisée par Prescott (cf [4]) permet aussi d'expliquer des s.e.u. expérimentaux sans maximum (si $t=1$, (2) s'écrit : $f(n) = \delta^n / (1+\delta)^{n+1}$, qui pour $n = 1, 2, \dots, \infty$ est une fonction décroissante de n). Néanmoins la loi $g(\delta')$ associée a l'inconvénient de donner des probabilités non négligeables à des gains δ' relativement élevés et physiquement impossibles.

On pourrait concevoir d'autres lois pour $g(\delta')$ que la loi gamma, par exemple la loi normale de moyenne μ et de variance σ^2 , mais il semble préférable d'essayer d'estimer directement la courbe $g(\delta')$ en écrivant $f(n)$ sous la forme d'une combinaison linéaire finie de p lois de Poisson de paramètre δ fixé, et en estimant les poids de chacune de ces p lois par des méthodes de régression. Cette approche qui a été effectuée avec p = 7 ou 8 lois de Poisson (cf [2]) a nécessité l'emploi de la régression sur variables entachées d'erreurs (cf [Régr.Contr.], Cahier, Vol III, n° 2, § 3.3, p 165) pour qu'on obtienne des résultats cohérents et physiquement interprétables. On obtient ainsi une estimation de la courbe continue $g(\delta')$ à l'aide d'un histogramme à 7 ou 8 classes. Cette estimation étant assez grossière, nous cherchons ici à obtenir une estimation plus précise de $g(\delta')$ à l'aide d'un histogramme ayant 30 à 50 classes. Pour effectuer cette estimation, on effectue l'analyse des correspondances des s.e.u. théoriques correspondant aux p lois de Poisson (p = 30 à 50) considérés sur le premier étage (les autres étages étant supposés poissonniens, et de gain constant), et on projette les s.e.u. expérimentaux étudiés en supplémentaire sur les premiers axes factoriels de cette analyse (cf Régr.Anal., § 5.2.1, ce cahier, pp 385-391).

2 Formulation statistique du problème physique et résolution

2.1 Formulation du problème

Soit $f_i(n)$ la distribution du gain du i^{ème} étage du P.M., et $G_i(s)$ la fonction génératrice associée :

$$G_i(s) = \sum \{s^n f_i(n) | n = 0, \infty\}$$

La fonction génératrice $G(s)$ du gain du P.M., i.e. du s.e.u. s'écrit,

si q est le nombre d'étages du P.M. .

$$G(s) = G_1(G_2(G_3 \dots (G_q(s)) \dots))$$

Posant

$$H(s) = G_2(G_3(\dots (G_q(s)) \dots))$$

on a :

$$G(s) = G_1(H(s))$$

Si $f_1(n)$ est la loi de Poisson de paramètre δ , on a $G_1(s) = \exp\{(s-1)\delta\}$ et donc $G(s) = \exp\{(H(s) - 1)\delta\}$.

Si $f_1(n)$ est fourni par la formule (1), ce que nous supposerons par la suite, on a :

$$G_1(s) = \int g(\delta') \exp\{(s-1)\delta'\} d\delta', \text{ et donc :}$$

$$G(s) = \int g(\delta') \exp\{(H(s)-1)\delta'\} d\delta' \\ = \int g(\delta') G(\delta', s) d\delta'$$

où $G(\delta', s)$ désigne la fonction génératrice du P.M. quand le premier étage est poissonnien avec un gain moyen égal à δ' .

Il en résulte que la distribution du gain du P.M. $p(n)$, i.e. le s.e.u. peut se mettre sous la forme :

$$p(n) = \int P(n, \delta') g(\delta') d\delta' \tag{3}$$

où $P(n, \delta')$ est la distribution du gain du P.M. quand le premier étage est poissonnien et de paramètre δ' , distribution que l'on peut considérer comme un s.e.u. théorique pur.

Si l'on suppose que les étages 2 à q sont poissonniens et de gain moyen δ'' donné, $P(n, \delta')$ est entièrement déterminé, et le problème est de trouver la fonction $g(\delta')$ telle que $P(n)$ qui est le s.e.u. théorique soit le plus proche possible du s.e.u. expérimental que nous désignerons par $y(n)$; adoptant la métrique du chi-deux de centre $y(n)$, métrique qui s'impose, on cherchera donc $g(\delta')$ de telle sorte que l'écart entre $y(n)$ et $P(n)$

$$\Sigma \{ (y(n) - p(n))^2 / y(n) \mid n = 1, \infty \} \tag{4}$$

soit minimum

Soit a la valeur maximale de δ' pour laquelle on est sûr que $g(\delta') = 0$ si δ' est supérieur ou égal à a . Divisant l'intervalle $(0-a)$ en p intervalles $I_1 = (a_0, a_1)$, $I_2 = (a_1, a_2)$, ..., $I_p = (a_{p-1}, a_p)$, avec $a_0 = 0$, $a_p = a$, (3) s'écrit approximativement en désignant par $c_i = (a_{i-1} + a_i)/2$ le centre de I_i :

$$P(n) = \Sigma \{ P(n, c_i) g(c_i) (a_i - a_{i-1}) \mid i = 1, p \}$$

Divisant l'intervalle $\{1, 2, 3, \dots, \infty\}$ en m intervalles consécutifs disjoints, J_1, \dots, J_m , on est ramené, en posant

$$\left. \begin{aligned} k(i, j) &= \Sigma \{ P(n, c_i) \mid n \in J_j \} \\ y_j &= \Sigma \{ y(n) \mid n \in J_j \} \\ b_i &= g(c_i) (a_i - a_{i-1}) \end{aligned} \right\} \tag{5}$$

à rechercher les b_i tels que :

$$\Sigma\{(y_j - \Sigma\{b_i k(i,j) \mid i = 1,p\})^2 / y_j \mid j = 1,m\} \quad (6)$$

soit minimum, ce qui revient à projeter avec la métrique adoptée le vecteur y_J des y_j sur le sous-espace de R_J engendré par les vecteurs k_{iJ} ($1 \leq i \leq p$) de composantes k_{ij} ($1 \leq j \leq m$).

En fait, pour que la solution obtenue ait un sens, les b_i doivent être positifs ou nuls ; on projettera donc y_J sur le cône positif du sous-espace précédent.

Par ailleurs, au lieu de minimiser (6), on minimisera la quantité:

$$\Sigma\{(y_j - \Sigma\{b_i k(i,j) \mid i = 1,p\})^2 / k(j) \mid j = 1,m\} \quad (7)$$

$$\text{où } k(j) = \Sigma\{k(i,j) \mid i = 1,p\} \quad (8)$$

ce qui revient à adopter comme métrique dans R_J la métrique du chi-deux de centre $f_J = k_J/k$ si k désigne la somme des $k(j)$, i.e. le total des éléments du tableau k_{IJ} des $k(i,j)$.

Remarques

1) Les b_i doivent être non seulement positifs, mais de somme inférieure ou égale à 1. Nous n'avons pas jugé bon d'imposer cette contrainte supplémentaire, la somme des b_i étant dans tous les essais que nous avons effectués, très voisine de 1, et toujours inférieure à 1,006.

2) Pour calculer les s.e.u. purs, i.e. les $P(n, \delta')$ et donc les $k(i,j)$, nous avons utilisé le résultat suivant (cf [3]) :

Soit X une variable aléatoire à valeurs dans N et de fonction génératrice $A(s) = \Sigma\{a_n s^n \mid n = 0, \infty\}$, où $a_n = \Pr(X=n)$; soit Y une variable aléatoire suivant une loi de Poisson de paramètre λ , et donc de fonction génératrice $B(s) = \exp\{(s-1)\lambda\}$; alors la variable aléatoire Z de fonction génératrice :

$$C(s) = \Sigma\{c_n s^n \mid n = 0, \infty\} = B(A(s)) = \exp\{(A(s)-1)\lambda\}$$

est telle que :

$$\left. \begin{aligned} c_0 &= \Pr(Z=0) = \exp\{(a_0-1)\lambda\} \\ \forall n \neq 0 : c_n &= \Pr(Z=n) = (\lambda/n) \Sigma\{(n-i) c_i a_{n-i} \mid i = 0, n-1\} \end{aligned} \right\} \text{(8bis)}$$

formules permettant de calculer rapidement les coefficients c_n de $C(s)$ connaissant $A(s)$ et λ .

Utilisant les formules précédentes, on calcule successivement $G_{q-1}(G_q(s))$ puis $G_{q-2}(G_{q-1}(G_q(s)))$, etc, d'où les $P(n, \delta')$ et les $k(i,j)^*$

2.2 Etude géométrique

Pour résoudre le problème précédent, on adopte un découpage de l'intervalle $0 - a$ ($a \leq 19$) des gains possibles en une quarantaine d'intervalles, découpage qui semble assez fin pour le problème étudié, et l'on effectue l'analyse factorielle des correspondances du tableau k_{IJ} associé en plaçant y_J en supplémentaire.

* Le programme de calcul des $k(i,j)$ que nous avons utilisé et qui est basé sur les formules (8bis), a été écrit par J.P. Ballini.

Dans le plan des axes 1-2, les points i se placent sur une parabole (effet Guttman), et pour que y_J soit une combinaison linéaire des k_{iJ} (combinaison linéaire à coefficients positifs et de somme inférieure ou égale à 1), il doit se projeter dans la concavité de cette parabole (cf fig 1a)

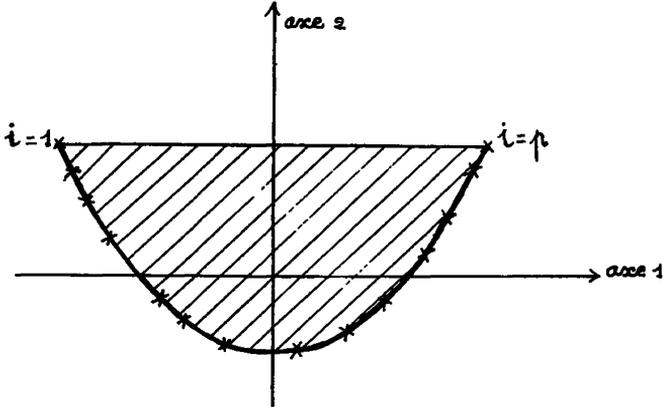


Figure 1a : Dans l'espace des deux premiers facteurs les points i projections des s.e.u. explicatifs se trouvent sur une parabole. Toute combinaison linéaire à coefficients positifs et de somme inférieure ou égale à 1 de ces s.e.u. doit se trouver à l'intérieur du domaine convexe hachuré.

De façon précise, à la loi $g(\delta')$ traduisant la combinaison linéaire des k_{iJ} associée à la loi de y_J , correspond dans le plan des axes 1-2 une loi ayant pour support la parabole précédente ; si $\alpha x^2 + \beta x + \gamma$ désigne l'équation de cette parabole, les coordonnées (x_1, x_2) du centre de gravité de cette loi sont données par :

$$\left. \begin{aligned} x_1 &= \mu_1 \\ x_2 &= \alpha(\mu_1^2 + \sigma_1^2) + \beta \mu_1 + \gamma \end{aligned} \right\} (9)$$

μ_1 et σ_1^2 désignant moyenne et écart-type de la loi sur le premier facteur.

Si $F_1(y)$ et $F_2(y)$ désignent les coordonnées de y_J sur les axes factoriels 1 et 2, les estimateurs $\hat{\mu}_1$ et $\hat{\sigma}_1^2$ de μ_1 et σ_1^2 sont obtenus en écrivant que la projection de y_J sur le plan des axes 1-2 est au centre de gravité de la parabole, ce qui revient à résoudre les équations (9) où l'on a remplacé x_1, x_2, μ_1 et σ_1^2 par $F_1(y), F_2(y), \hat{\mu}_1$, et $\hat{\sigma}_1^2$; l'on obtient ainsi :

$$\left. \begin{aligned} \hat{\mu}_1 &= F_1(y) \\ \hat{\sigma}_1^2 &= (F_2(y) - \beta F_1(y) - \gamma) / \alpha - (F_1(y))^2 \end{aligned} \right\} (10)$$

De même dans le plan 1-3, les points i se placent sur une cubique, tandis que dans l'espace des trois premiers facteurs, ils se trouvent

sur une cubique gauche C intersection de deux cônes ayant une génératrice commune, la droite $A_1 A_p$ joignant les points i extrêmes ($i = 1$ et $i = p$) ; chacun de ces cônes s'appuie donc sur C , le premier ayant pour sommet A_1 , et le second pour sommet A_p (cf fig. 1b). Pour que y_J soit une combinaison linéaire des k_{iJ} (ou ce qui est équivalent des profils f_J^i des lignes du tableau k_{IJ}) sa projection dans l'espace des trois premiers axes factoriels doit se trouver à l'intérieur du convexe délimité

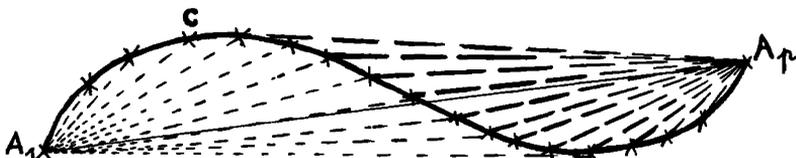


Figure 1b : Dans l'espace des trois premiers facteurs les points A_i ($1 \leq i \leq p$) représentatifs des s.e.u. explicatifs se projettent suivant une cubique gauche C . Toute combinaison linéaire à coefficients positifs et de somme inférieure ou égale à 1 de ces s.e.u. doit se trouver dans le convexe délimité par deux cônes ayant une génératrice commune, la droite $A_1 A_p$ joignant les points extrêmes de C . Ces deux cônes s'appuient sur C , le premier ayant pour sommet A_1 , et le second A_p .

par les deux cônes précédents (convexe qui contient l'origine, puisque les facteurs sont centrés).

Dans l'espace des r premiers facteurs ($r \geq 4$), la projection z_J^r de y_J doit encore se trouver dans le convexe C_r formé par les combinaisons linéaires à coefficients positifs et de somme inférieure ou égale à 1 des projections des points i , mais ce convexe n'a plus une signification géométrique aussi simple que dans le cas où $r = 2$ ou 3. On peut s'assurer que cette projection se trouve à l'intérieur de ce convexe, si on peut trouver dans cet espace une combinaison linéaire à coefficients positifs et de somme inférieure ou égale à 1 des projections des points i reconstituant exactement la projection de y_J .

D'un point de vue pratique, pour trouver cette combinaison linéaire, il suffit de se servir d'un programme de régression sous contraintes de positivité des coefficients de régression, programme où l'on impose de plus à la somme des coefficients de régression d'être inférieure ou égale à 1. Si l'on attaque ce programme avec un vecteur de régression initial satisfaisant aux contraintes (i.e. en un point intérieur à C_r) et si z_J^r se trouve à l'intérieur de C_r , on obtient un ajustement parfait entre z_J^r et son approximation zz_J^r donnée par le programme, les coefficients de la combinaison linéaire ainsi trouvée dépendant bien sûr du vecteur de régression initial, puisque dans cette régression sous contraintes, on a un nombre de variables explicatives $\text{Card } I = p$ supérieur à la dimension r de l'espace où l'on se trouve. Si par contre z_J^r est à l'extérieur de C_r , l'ajustement entre z_J^r et zz_J^r ne sera plus parfait (e.g. $\|z_J^r - zz_J^r\|^2 / \|z_J^r\|^2 \geq 10^{-3}$).

On a ainsi un moyen de déterminer le nombre r_0 de facteurs à conserver. On appelle le programme précédent avec 2 facteurs, 3 facteurs, etc, et si l'on obtient un ajustement parfait avec h facteurs, et un ajustement non parfait avec $h+1$ facteurs, on conservera $r_0 = h$ facteurs, l'explication de y_j en fonction des k_{iJ} devenant illusoire si l'on garde plus de r_0 facteurs puisqu'une véritable amélioration ne peut se faire sans introduire de coefficients négatifs dépourvus de sens physique.

2.3 Estimation de $g(\delta')$

Pour estimer la loi $g(\delta')$, i.e. déterminer les coefficients b_i donnés en (5), au lieu de considérer le tableau k_{iJ} et de minimiser (6), on se place dans l'espace des r premiers axes factoriels de k_{iJ} (cf [Regr. Anal.], ce cahier pp 385-391).

Soit $(F_0, G_0), (F_1, G_1), \dots, (F_r, G_r)$ les $r+1$ premiers couples de facteurs associés issus de k_{iJ} , (F_α, G_α) ($0 \leq \alpha \leq r$) étant de variance λ_α valeur propre associée à ce couple, (F_0, G_0) désignant le couple de facteurs triviaux constants ($F_0(i) = G_0(j) = 1$) associé à la valeur propre $\lambda_0 = 1$. Si $u_{0J}, u_{1J}, \dots, u_{rJ}$ désignent les axes factoriels associés dans R_J ($u_{\alpha j} = (G_\alpha(j) / \sqrt{\lambda_\alpha}) k(j)/k$, raisonner dans l'espace des $r+1$ premiers facteurs ($r+1$ à cause de l'axe factoriel trivial) revient à remplacer les variables initiales par les $u_{\alpha J}$ ($0 \leq \alpha \leq r$) qui sont orthonormés (pour la métrique du chi-deux de centre $f_J = k_J/k$).

On recherche alors les coefficients c_α ($0 \leq \alpha \leq r$) minimisant :

$$\Sigma \{ (y_j - \Sigma \{ c_\alpha u_{\alpha j} \mid \alpha = 0, r \})^2 / k(j) \mid j = 1, m \} \tag{11}$$

Compte tenu de ce que :

$$u_{\alpha j} = k(j) G_\alpha(j) / (k \sqrt{\lambda_\alpha}) = \Sigma \{ k(i, j) F_\alpha(i) \mid i = 1, p \} / (k \lambda_\alpha) \tag{12}$$

l'approximation

$\Sigma y_j = \Sigma \{ c_\alpha u_{\alpha j} \mid \alpha = 0, r \}$ de y_j réalisée à l'aide des $r+1$ premiers axes factoriels, se met sous la forme : $\Sigma \{ b_i k(i, j) \mid i = 1, p \}$ avec :

$$\left. \begin{aligned} b_i &= \Sigma \{ c_\alpha F_\alpha(i) / (k \lambda_\alpha) \mid \alpha = 0, r \} \\ &= (c_0/k) + \Sigma \{ c_\alpha F_\alpha(i) / (k \lambda_\alpha) \mid \alpha = 1, r \} \end{aligned} \right\} \tag{13}$$

et pour que la solution obtenue ait un sens, b_i doit être positif ou nul.

On est donc amené à minimiser (11) sous les contraintes :

$$\forall i = 1, \dots, p : \Sigma \{ c_\alpha F_\alpha(i) / \lambda_\alpha \mid \alpha = 0, r \} \geq 0 \tag{14}$$

Notons que si l'on n'impose pas les contraintes (14) et si $F_\alpha(y)$ désigne la valeur du $\alpha^{\text{ème}}$ facteur F_α pour y_j placé en supplémentaire du tableau k_{iJ} , $F_\alpha(y)$ est l'abscisse de la projection du profil y_j/y_{tot} (si y_{tot} désigne la somme des y_j) sur $u_{\alpha J}$. On a donc alors : $c_\alpha = y_{\text{tot}} F_\alpha(y)$, et b_i s'écrit :

$$b_i = (y_{\text{tot}}/k) (1 + \Sigma \{ F_\alpha(y) F_\alpha(i) / \lambda_\alpha \mid \alpha = 1, r \}) \tag{15}$$

Remarques

1) y_j et les $k_{i,j}$ étant des lois de probabilité, on a $y_{\text{tot}} = k(i) = 1$ (où $k(i) = \sum \{k(i,j) \mid j = 1, m\}$) et donc $k = \text{Card } I = p$; le facteur y_{tot}/k dans la formule précédente est donc égale à $1/p$. Nous préférons garder l'expression générale (15) dans la mesure où elle est valable même si y_j et les $k_{i,j}$ ne sont plus des lois de probabilité. En particulier, les s.e.u. expérimentaux y_j étudiés au § 3 (et donc les s.e.u. théoriques purs $k_{i,j}$) ayant été tronqués de leurs cinq premières classes ne sont plus de masse totale 1, à moins qu'on ne les y ramène par homothétie (cf § 3.2 *in fine*, et début du § 3.4).

2) Au lieu de se placer dans l'espace des r premiers facteurs d.l'analyse des correspondances de $k_{i,j}$, on peut aussi se placer dans l'espace des r premiers facteurs de l'analyse en composantes principales (non centrée) du tableau $k_{i,j}$, R_I étant muni de la métrique de matrice, la matrice diagonale de $i^{\text{ème}}$ terme diagonal v_{ii}^{-1} si $v_{ii} = \sum \{k(i,j)k(i',j) \mid j = 1, m\}$, et R_J de la métrique usuelle (ou de la métrique du chi-deux de centre y_j). On obtient ainsi, si l'on n'impose pas des contraintes du type (14), l'estimateur de Marquardt (cf Méthodes de Régression, II Critères Bayésiens, cahiers, Vol III, n° 3, § 4.7, p. 264).

Si l'on impose ces contraintes, on obtient ainsi une estimation de $g(\delta')$ qui a un sens. Néanmoins, ayant affaire à des lois de probabilité, l'analyse des correspondances semble s'imposer ici.

D'un point de vue pratique, ayant effectué l'analyse des correspondances de $k_{i,j}$ avec y_j en supplémentaire, on a immédiatement d'après (15) les résultats de la régression quand on n'impose pas les contraintes (14). Pour trouver la solution avec contraintes, on opère pas à pas : au pas 0, avec le facteur trivial, tous les coefficients b_i fournis par (15), où $r = 0$ sont positifs et donc satisfont les contraintes. Si au pas $h-1$ ($r = h-1$) tous les coefficients fournis par (15) sont positifs, et si ce n'est pas le cas au pas h , on diminue en valeur absolue le coefficient $c_h = y_{\text{tot}} F_h(y)$ jusqu'à ce que tous les coefficients b_i fournis par (13) (où $c_\alpha = y_{\text{tot}} F_\alpha(y)$, pour $\alpha \leq h-1$) soient positifs ou nuls, et avec comme point de départ le vecteur des c_α ($0 \leq \alpha \leq h$) ainsi obtenu, vecteur satisfaisant (14), on attaque un sous-programme de programmation quadratique minimisant (11) sous les contraintes (14). Si $c_\alpha^{(h)}$ ($0 \leq \alpha \leq h$) désigne la solution du sous-programme précédent, et si au pas $h+1$, tous les coefficients fournis par (15) où $r = h+1$ ne sont pas positifs ou nuls, on attaque le sous-programme précédent avec comme point de départ le vecteur des c_α ($0 \leq \alpha \leq h+1$) défini ainsi : $c_\alpha = c_\alpha^{(h)}$ pour $0 \leq \alpha \leq h$; $c_{h+1} = \epsilon y_{\text{tot}} F_{h+1}(y)$, où ϵ est le plus grand nombre compris entre 0 et 1, et tel que (14) où $r = h+1$ soit vérifiée, un tel nombre existant toujours, puisque si $\epsilon = 0$, (14) est vérifiée par hypothèse.

Pour réaliser le minimum de (11) sous les contraintes (14), nous avons adapté notre sous-programme ESOPN de régression sous contrainte de positivité (cf Régr. Contr., cahiers Vol III n° 2 §§ 3.2.2, 3.2.5), sous-programme qui minimise (11) sous les contraintes $c_\alpha \geq 0$; le sous-programme

ainsi obtenu (*), nettement plus complexe, puisqu'il peut y avoir des pivotages (ce qui n'est pas le cas avec ESOPN) n'est pas optimal, mais dans la mesure où l'on part d'un point initial proche de la solution, il converge en quelques itérations (en général en moins de dix itérations).

3 Explication de 10 s.e.u. expérimentaux par la loi du gain du premier étage du P.M.

3.1 s.e.u. expérimental et s.e.u. normalisé

D'un point de vue expérimental, chaque fois qu'il arrive un paquet (**) de n électrons à l'anode du photomultiplicateur étudié, on obtient une impulsion de tension, qui est amplifiée, puis envoyée dans un analyseur multicanaux, où elle est enregistrée dans un canal associé à une tension V d'autant plus grande que n est élevé. A la fin de l'expérience on dispose donc de la courbe donnant en fonction de la tension V_i

($1 \leq i \leq t$, si t est le nombre de canaux de l'analyseur) une quantité $N(V_i)$ qui est proportionnelle au nombre de fois où un paquet d'électrons a donné lieu à une tension égale à V_i (i.e. à une tension comprise entre $V_i - \Delta V_i$ et $V_i + \Delta V_i$, si on a un canal de largeur $2\Delta V_i$ centré sur V_i , ce qui correspond à un nombre d'électrons compris entre deux quantités $n_i - \Delta n_i$ et $n_i + \Delta n_i$).

Les photoélectrons donnant naissance aux paquets d'électrons précédents, étant assez espacés, de par les conditions de l'expérience pour que ces paquets d'électrons n'interfèrent pas, la courbe obtenue $N(V_i)$ en fonction de V_i est le spectre à électrons unique (s.e.u.) expérimental.

Pour obtenir un s.e.u. normalisé (i.e. l'histogramme expérimental donnant pour un nombre d'électrons, (ou pour une classe (n_1, n_2)) la fréquence (relative) d'apparition de ce nombre, histogramme correspondant à la donnée du vecteur y_j , cf § 2.1), il faut d'abord normaliser l'échelle des abscisses (***), puis celle des ordonnées, cette dernière normalisation étant faite pour se ramener à une loi de probabilité. Pour normaliser l'échelle des abscisses, il suffit d'écrire que la valeur moyenne du s.e.u. expérimental est égal au gain moyen G du P.M., gain égal à $\delta \delta^{q-1}$ si le P.M. comporte q étages, le premier de gain δ et les autres de gain δ , δ et δ étant connus. Pour normaliser l'échelle des ordonnées, il suffit de ramener la surface comprise entre la courbe et l'axe des abscisses à avoir une valeur égale à 1, en divisant chaque ordonnée par cette surface.

(*) Ce sous-programme peut également minimiser une expression du type (11) avec les contraintes $c_\alpha \geq 0$, $\sum c_\alpha \leq 1$, ce qui permet (cf 2.2 in fine) de déterminer le nombre de facteurs à conserver pour faire la régression.

(**) Paquet obtenu à partir d'un photoélectron issu de la photocathode, i.e. d'un électron libéré par un photon tombant sur la photocathode d'entrée du P.M. .

(***) i.e. avoir la correspondance entre le nombre d'électrons n à la sortie du P.M., et la tension V à laquelle ils donnent naissance, V étant proportionnel à n : $V = \lambda n$.

Une difficulté se présente dans les normalisations précédentes, pour calculer la valeur moyenne du s.e.u. expérimental, et la surface comprise entre ce s.e.u. et l'axe des abscisses : on ne connaît pas la valeur $N(0)$, i.e. le nombre de fois où un photoélectron issu de la photocathode du P.M. ne donne naissance à aucun électron à la sortie du P.M.. Pour lever cette difficulté, on peut opérer les normalisations en considérant les spectres tronqués de l'origine, ce qui revient à considérer pour le P.M., un gain égal à $G/(1 - P(0))$, $P(0)$ étant la probabilité théorique qu'un photoélectron ne donne naissance à aucun électron à la sortie du P.M.. La distribution du premier étage n'étant pas connue, puisqu'on veut l'estimer, $P(0)$ n'est pas connu ; néanmoins, en première approximation, on pourra estimer $P(0)$ en supposant le premier étage poissonnien et de gain δ (les autres étages étant bien sûr toujours supposés poissonniens, et de gain δ'').

Une autre voie, qui a été celle adoptée par le physicien est de donner à $P(0)$ une valeur p (ce qui revient encore à se donner $N(0)$), d'en déduire le s.e.u. normalisé (non tronqué), et d'estimer la distribution du gain du premier étage ; à partir de cette distribution, on peut recalculer la valeur de $P(0)$; soit p' la valeur ainsi calculée ; si p' est différent de p , on recommence le processus précédent avec comme valeur initiale de $P(0)$, la valeur p' ainsi calculée, et on réitère le processus jusqu'à ce qu'il converge, ce qui, pour les s.e.u. analysés, avait lieu en 2 ou 3 itérations.

3.2 Les données

Les données en notre possession correspondent à 10 spectres à électron unique normalisés, obtenus à partir de 10 spectres à électron unique expérimentaux enregistrés dans des conditions d'expériences identiques (même P.M., mêmes conditions d'éclairement et de voltage etc).

Il semble en fait qu'entre ces dix spectres, il y ait des différences un peu plus importantes que celles dues aux fluctuations de comptage, différences pouvant s'expliquer par des variations erratiques des conditions de mesure. Néanmoins ces dix spectres représentant le même phénomène physique, l'estimation des fonctions $g(\delta')$ associées permet de se rendre compte de la variabilité de ces estimations.

Chacun des 10 s.e.u. expérimentaux précédents a été normalisé suivant la seconde voie indiquée au § 3.1 *in fine*, le gain moyen du premier étage qui est compris entre 7 et 8,5 suivant les conditions d'éclairement de la photocathode ayant été pris égal à 7,65, et le gain des étages suivants (en nombre égal à 13) à 3,075. Avec une estimation de la distribution du gain du premier étage à l'aide d'un histogramme à 7 ou 8 classes (cf [2]), on obtient des $P(0)$ faibles et de l'ordre de 1 à 3%.

Les s.e.u. normalisés ainsi obtenus comportent 100 classes ; en fait, on s'est contenté d'expliquer les 95 dernières classes, car les 5 premières, risquant d'un point de vue expérimental d'être perturbées par des phénomènes parasites, ont été annulées. Notons qu'il aurait mieux valu, puisqu'on n'essaie pas d'expliquer les premières classes des s.e.u., normaliser les s.e.u. suivant la première voie indiquée au § précédent.

Par ailleurs chaque s.e.u. expérimental a été lissé avant d'être normalisé, alors qu'il aurait mieux valu garder les s.e.u. expérimentaux initiaux.

Ne disposant que des s.e.u. normalisés comme il est indiqué ci-dessus, avec 95 classes (on ne dispose plus des s.e.u. expérimentaux initiaux, ni des 5 premières classes des s.e.u. normalisés) ce sont sur ces 10 s.e.u. que nous avons travaillé.

3.3 Les découpages adoptés

Nous avons considéré plusieurs découpages de l'intervalle (0-20) des gains possibles pour le premier étage du P.M. étudié. Le premier découpage considéré correspond à 40 intervalles de centre, les valeurs c_i suivantes :

0,25 ; 0,50 ; 0,75 ; 1 ; 1,25 ; 1,50 ; 1,75 ; 2 ; 2,33 ; 2,67 ; 3 ; 3,33 ; 3,67 ; 4 ; 4,33 ; 4,67 ; 5 ; 5,50 ; 6 ; 6,50 ; 7 ; 7,50 ; 8 ; 8,50 ; 9 ; 9,50 ; 10 ; 10,5 ; 11 ; 11,5 ; 12 ; 12,5 ; 13 ; 13,5 ; 14 ; 15 ; 16 ; 17 ; 18 ; 19 ;

ce qui revient, rappelons-le, à faire l'hypothèse que la statistique de multiplication du premier étage est une combinaison linéaire de lois de Poisson de valeur moyenne les valeurs c_1 précédentes.

Après avoir effectué diverses analyses des correspondances - analyse des correspondances du tableau k_{IJ} des s.e.u. théoriques associés au découpage précédent (cf § 2.1, *in fine*), avec Card I = 40, Card J = 95, et analyses de sous-tableaux $k_{I',J}$ de k_{IJ} ($I' = I - I'$ étant mis en supplémentaire), sous-tableaux où tout en gardant le même découpage que précédemment, l'on considère une plage plus restreinte que la plage 0,25 - 19 - nous nous sommes limités à la plage 1,5 - 17 du découpage précédent, ce qui correspond à un découpage en 33 classes. Si I' désigne le sous-ensemble de I associé à ces 33 classes, l'ensemble I' dessine dans le plan 1-2 de l'analyse factorielle de $k_{I',J}$, une parabole d'axe le second axe factoriel (cf figure 2), tandis que les 10 s.e.u. expérimentaux (après normalisation) (*), s.e.u. notés Y_1, Y_2, \dots, Y_{10} et qui se projettent pratiquement au même endroit sur ce plan, ont leur projection au voisinage de cet axe (à l'intérieur de la concavité de la parabole), ce qui correspond à un bon "centrage" de ces s.e.u. par rapport aux s.e.u. théoriques (i.e. à l'ensemble I') dont ils sont censés être une combinaison linéaire.

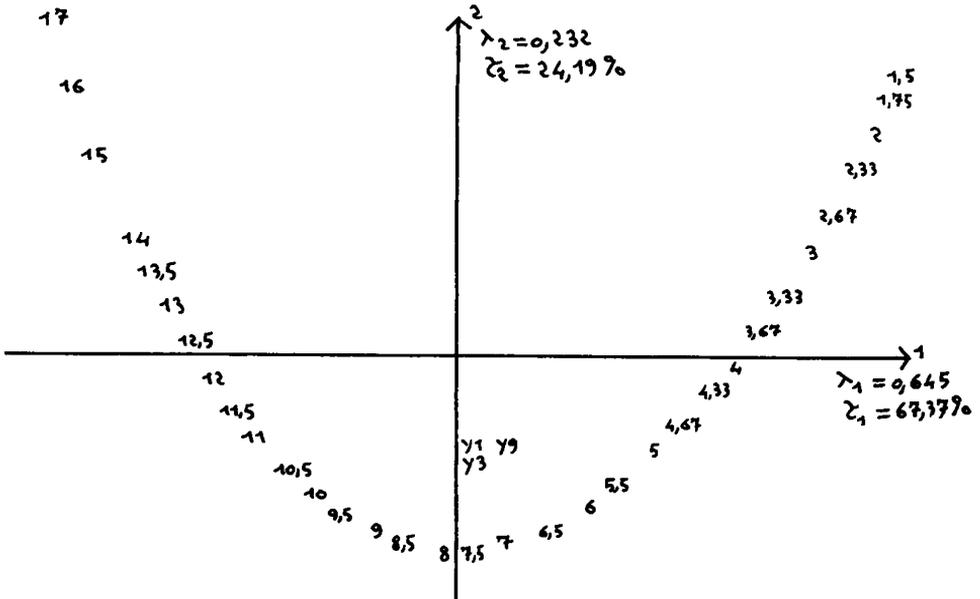


Figure 2 : Plan 1-2 de l'analyse factorielle du tableau $(33 \times 95) k_{I',J}$ associé à un découpage en 33 classes de la plage 1,5-17. Les s.e.u. expérimentaux Y_2, Y_4 à Y_8, Y_{10} sont confondus avec Y_1 .

(*) Nous parlerons par la suite de s.e.u. expérimentaux, pour désigner les s.e.u. expérimentaux normalisés.

De plus la plage 1,5 - 17 correspond (et ceci pour chacun des 10 s.e.u. expérimentaux) à un intervalle légèrement plus grand que l'intervalle $(\hat{\mu}_1 - 2\hat{\sigma}_1, \hat{\mu}_1 + 2\hat{\sigma}_1)$ (cf tableau 1) centré sur $\hat{\mu}_1$, et de longueur $4\hat{\sigma}_1$, $(\hat{\mu}_1, \hat{\sigma}_1)$ désignant respectivement les estimations de la moyenne et de l'écart-type de la loi du premier facteur associé au s.e.u. expérimental y considéré (cf § 2.2) estimations données par les formules (10). Dans le cas où l'on a une loi normale, cette plage contient donc plus de 95% de la masse de la loi de y.

n° du s.e.u.	1	2	3	4	5	6	7	8	9	10
$10^4 \hat{\mu}_1$	-195	-84	-7	-43	-184	-146	-107	232	303	251
$\hat{\sigma}_1$	0,586	0,579	0,575	0,592	0,593	0,606	0,595	0,587	0,583	0,581
$\hat{\delta}$	7,72	7,67	7,63	7,65	7,72	7,70	7,58	7,52	7,49	7,51

Tableau n° 1 : Moyenne $\hat{\mu}_1$ et écart-type $\hat{\sigma}_1$ de la loi associée à chaque s.e.u. expérimental, et ayant pour support la parabole du plan 1-2 issu de l'analyse factorielle du tableau (33 x 95) $k_{I,J}$ correspondant à la plage 1,5-17. $\hat{\mu}_1$ égal à $F_1(y)$ étant compris soit entre $F_1(8) = -0,077$ et $F_1(7,50) = 0,028$, soit entre $F_1(7,50)$ et $F_1(7) = 0,136$, on a par interpolation déduit la valeur $\hat{\delta}$ du gain moyen du premier étage.

A l'aide d'un sous-programme de régression sous contraintes imposant aux coefficients d'être positifs et de somme inférieure ou égale à 1 (cf § 2.2 *in fine*, et § 2.3, renvoi p402), on a pu montrer que dans l'espace des $r=5$ premiers facteurs, les dix spectres expérimentaux (ou plus exactement leur projection sur cet espace (*)) sont à l'intérieur du convexe C_r délimité par les combinaisons linéaires à coefficients positifs et de somme inférieure ou égale à 1 des projections des s.e.u. théoriques i de I' conservés ; pour $r=6$ facteurs, les s.e.u. expérimentaux (*) sont à la frontière du convexe C_r , tandis que pour $r=7$, ils sont à l'extérieur. Pour estimer les coefficients b_i de la combinaison linéaire $\Sigma\{b_i k_{iJ} | i \in I'\}$ ajustant l'un des 10 s.e.u. expérimentaux, on se place donc dans l'espace des six premiers facteurs de l'analyse factorielle, espace correspondant à 99.997% de l'inertie (cf tableau 2).

On a vérifié que si l'on restreint la plage 1,5 - 17, la dimension r_0 de l'espace factoriel pour lequel les 10 spectres expérimentaux sont à l'intérieur du convexe C_{r_0} associé diminue. Ainsi avec la plage 3-13,5 correspondant à un découpage en 24 classes, $r_0 = 4$; avec la plage 4-12,5 (découpage en 19 classes), $r_0 = 3$, tandis qu'avec la plage 4,33-11,5 (découpage en 16 classes) $r_0 = 2$, r_0 étant égal à 1 pour la plage 5-10,5 associé à un découpage en 12 classes.

(*) Dans la suite, on dira souvent, par abus de langage qu'un s.e.u. appartient au convexe C_r associé aux r premiers facteurs pour signifier que la projection de ce s.e.u. sur l'espace des r premiers facteurs appartient à C_r .

axe	valeur propre	% d'inertie	% d'inertie cumulé
1	0,64535	67,368	67,368
2	0,23177	24,194	91,562
3	0,06559	6,847	98,409
4	0,01313	1,370	99,779
5	0,00187	0,195	99,974
6	0,00023	0,024	99,997

Tableau n° 2 : Valeurs propres et pourcentages d'inertie associés à l'analyse factorielle du tableau (33×95) $k_{I,J}$ correspondant à la plage 1,5-17.

Notons que les plages précédentes ont été choisies de façon à avoir dans le plan des axes 1-2 une parabole d'axe le second axe factoriel, et de façon à ce que les 10 s.e.u. expérimentaux se projettent au voisinage de ce second axe, ce qui revient à avoir pour projection des points 7, 7,5 et 8 sur le premier axe factoriel des points voisins de l'origine, puisque le gain moyen du premier étage du P.M. est de cet ordre de grandeur.

Si l'on estime pour chaque s.e.u. expérimental les coefficients b_i à l'aide du programme de régression sous contraintes mentionné au § 2.3, et en se plaçant dans l'espace des six premiers facteurs de $k_{I,J}$ (où I' correspond à la plage 1,5-17), on obtient une variation continue des coefficients b_i en fonction des centres c_i des intervalles associés au découpage adopté. Par contre, si l'on considère l'histogramme correspondant, i.e. la courbe donnant $g(c_i)$ en fonction de c_i (cf formule (5)), on obtient des discontinuités en début de spectre (i.e. pour les valeurs de c_i faibles) discontinuités dues au fait que l'on n'a pas un pas constant dans le découpage, et survenant quand on a un changement de pas, i.e. pour les valeurs de c_i égales à 2 et 5.

Pour cette raison, on a considéré un découpage de l'intervalle 0-20 en intervalles de même longueur 0,5, et après différentes analyses factorielles du tableau $k_{I,J}$ associé, et de sous-tableaux de $k_{I,J}$, on a retenu le découpage suivant :

0,5 ; 1 ; 1,5 ; 2 ; 2,5 ; 3 ; 3,5 ; 4 ; 4,5 ; 5 ; 5,5 ; 6 ; 6,5 ; 7 ; 7,5 ; 8 ; 8,5 ; 9 ; 9,5 ; 10 ; 10,5 ; 11 ; 11,5 ; 12 ; 12,5 ; 13 ; 13,5 ; 14 ; 15 ; 16 ; 17 ;

découpage en 31 classes, où les valeurs précédentes sont associées aux centres des classes considérées. Ce découpage correspond à un pas égal à 0,5, entre 0,5 et 14, puis égal à 1 entre 14 et 17.

Il a été choisi de telle sorte que comme précédemment, on obtienne dans le plan 1-2 de l'analyse des correspondances du tableau associé une parabole d'axe, le deuxième axe factoriel, les 10 s.e.u. expérimentaux se projetant au voisinage de cet axe. On a préféré débiter le découpage par 0,5, des gains faibles de l'ordre de 0,5 ou 1 étant physiquement possibles.

C'est donc sur ce découpage que l'on raisonnera par la suite. On désignera par I l'ensemble des classes : $I = \{0,5 ; 1 ; 1,5 ; \dots ; 13,5 ; 14 ; 15 ; 16 ; 17\}$ et par $k_{I,J}$ le tableau des s.e.u. théoriques associés. On continuera par appeler I' le sous-ensemble $\{1,5 ; 1,75 ; \dots, 17\}$ associé au

premier découpage (1,5-17) considéré, et par k_{IJ} le tableau des s.c.u. correspondants.

Les résultats de l'analyse factorielle du tableau k_{IJ} (ou plutôt du tableau k'_{IJ} , où l'on a ramené les s.e.u. théoriques k'_{iJ} à avoir masse totale 1, cf § 3.4) sont reportés sur le tableau 3 et la figure 3 où sont représentés les plans 1-2 et 1-3.

axe	valeur propre	% d'inertie	% d'inertie cumulé
1	0,6857	62,44	62,44
2	0,2923	26,61	89,05
3	0,0935	8,52	97,57
4	0,0222	2,02	99,59
5	0,0039	0,35	99,94
6	0,0005	0,05	99,99
7	0,00006	0,006	100

Tableau n° 3 : Valeurs propres et pourcentages d'inertie associés à l'analyse factorielle du tableau (31 × 95) k'_{IJ} correspondant à la plage 0,5-17.

3.4 Estimation de la loi $g(\delta')$ du gain du premier étage du P.M.

Ne cherchant à expliquer que les 95 dernières classes des s.e.u. expérimentaux normalisés (cf § 3.2) puisqu'ils ont été tronqués de leurs 5 premières classes, les 31 s.e.u. théoriques associés au découpage I considéré ont également été tronqués de leurs cinq premières classes. Il en résulte que si J désigne l'ensemble des 95 classes des s.e.u. théoriques et expérimentaux, et si y_J et k_{iJ} désignent respectivement l'un des dix s.e.u. expérimentaux, et le $i^{\text{ème}}$ s.e.u. théorique, on a :

$$\Sigma \{y_j \mid j \in J\} = y_{\text{tot}} < 1$$

$$\Sigma \{k(i,j) \mid j \in J\} = k(i) < 1$$

Au lieu de chercher à expliquer y_J par la combinaison linéaire $\Sigma \{b'_i k_{iJ} \mid i \in I\}$, on peut chercher à expliquer $z_J = y_J / y_{\text{tot}}$ par la combinaison linéaire $\Sigma \{b'_i k_{iJ} / k(i) \mid i \in I\} = \Sigma \{b'_i k'_{iJ} \mid i \in I\}$ (où $k'_{iJ} = k_{iJ} / k(i)$), ce qui revient à raisonner sur des lois de probabilité (i.e. lois de masse totale 1). Dans ce cas on fera l'analyse factorielle du tableau k'_{IJ} des $k(i,j) / k(i)$. A partir des coefficients b'_i ainsi estimés, on en déduit les coefficients b_i par la formule :

$$b_i = b'_i y_{\text{tot}} / k(i) \quad (16)$$

Si l'on se plaçait dans R_J tout entier, les deux approches précédentes (analyse des correspondances de k_{IJ} , ou analyse des correspondances de k'_{IJ}) conduiraient aux mêmes coefficients b_i . Comme l'on se place dans l'espace engendré par les premiers facteurs de ces analyses, on obtient en fait des résultats légèrement différents. Nous avons expérimenté ces deux approches ; comme elles donnent des résultats sensiblement équivalents, nous ne donnerons par la suite, sauf mention contraire, que ceux obtenus à partir de l'analyse du tableau k'_{IJ} .

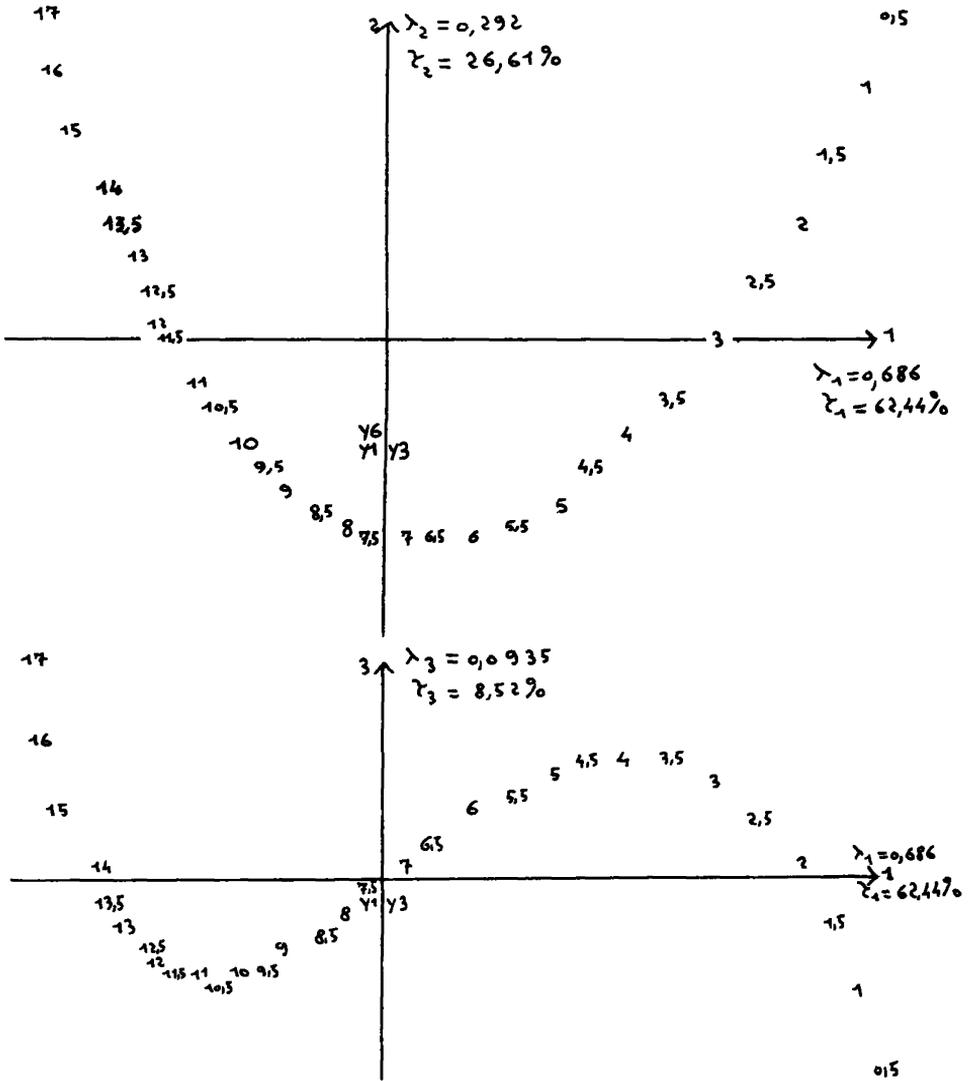


Figure 3 : Plans 1-2 et 1-3 de l'analyse factorielle du tableau $(31 \times 95) k'_{IJ}$ associé à un découpage en 31 classes de la plage 0,5-17. Sur ces deux plans les s.e.u. expérimentaux Y4, Y7 à Y10 sont confondus avec Y3, tandis que Y2 et Y5 sont confondus avec Y1. Sur le plan 1-3, Y6 est également confondu avec Y1.

Les 10 s.e.u. expérimentaux sont à l'intérieur (ou sur la frontière) du convexe C_7 associé aux 7 premiers axes factoriels de l'analyse des correspondances de k'_{IJ} , et à l'extérieur du convexe C_8 associé aux 8 premiers facteurs.

On se placera donc dans l'espace des 7 premiers facteurs pour estimer les coefficients b'_i .

On peut noter que la régression usuelle (i.e. sans imposer les contraintes (14)) donne des résultats pratiquement acceptables si l'on se place dans l'espace des 7 premiers facteurs, des coefficients de régression b'_i négatifs s'introduisant certes, mais à l'extrémité supérieure du découpage I, et avec une valeur absolue faible. Par contre, si l'on se place dans l'espace des huit premiers facteurs, on obtient des résultats aberrants, avec des coefficients de régression b'_i négatifs et non négligeables. Ceci est bien évidemment en accord avec le fait que les s.e.u. expérimentaux sont à l'extérieur du convexe C_8 .

Par contre si l'on impose les contraintes (14), on obtient le même histogramme des b'_i que l'on raisonne dans l'espace des 7 ou 8 (ou même 9) premiers facteurs. En effet, le fait d'imposer ces contraintes entraîne la nullité des coefficients de régression associés au 8^o (et 9^o) facteur. On retrouve là l'effet bien connu de protection de la régression quand on impose des contraintes (cf [Régr. Contr.], Cahiers, Vol III, n^o 2 § 2).

Les estimations de l'histogramme des b'_i relative aux 10 s.e.u. expérimentaux étant très voisines, on a calculé pour chaque valeur de i ($1 \leq i \leq 31$) la moyenne \bar{b}'_i des dix b'_i associés ainsi que l'écart-type empirique correspondant σ_i et l'on a reporté sur la figure n^o 4 l'histogramme $g'(\delta_i)$ des \bar{b}'_i relatifs à l'histogramme moyen (i.e. avec le découpage adopté : $g'(\delta_i) = 2 \bar{b}'_i$ pour $0,5 \leq \delta_i \leq 13,5$; $g'(\delta_i) = (4/3) \bar{b}'_i$ si $\delta_i = 14$; $g'(\delta_i) = \bar{b}'_i$ si $\delta_i \geq 15$). On a encadré la courbe précédente des deux courbes $g''(\delta_i)$ et $g'''(\delta_i)$ obtenues en retranchant et ajoutant respectivement σ_i à \bar{b}'_i (i.e. $g''(\delta_i)$ est relatif aux $b''_i = \bar{b}'_i - \sigma_i$, tandis que $g'''(\delta_i)$ est relatif aux $b'''_i = \bar{b}'_i + \sigma_i$), afin de visualiser la dispersion des résultats.

A partir des b'_i , on peut calculer à l'aide de la formule (16) les b_i relatifs à chaque s.e.u. expérimental, et donc en déduire la moyenne \bar{b}_i associée ; on peut alors tracer l'histogramme des \bar{b}_i , qui d'un point de vue physique est ce qui nous intéresse.

A titre de comparaison, on a reporté sur la figure 5 l'histogramme des \bar{b}_i ainsi calculés (histogramme obtenu à partir de l'analyse factorielle de k'_{IJ}), ainsi que l'histogramme moyen des b_i calculés directement à partir de l'analyse factorielle de k_{IJ} .

On voit qu'à part les trois premières valeurs 0,5 ; 1 ; 1,5 ; (et essentiellement la première) du découpage, ces histogrammes sont pratiquement identiques. On peut noter que pour ces trois premières valeurs, on a une remontée de l'histogramme, remontée déjà notée dans le cas de l'histogramme à 7 ou 8 classes étudié précédemment (cf [2]). Néanmoins cette remontée n'est probablement pas significative, l'estimation du début et de la fin de l'histogramme étant relativement peu précise, comme l'on a pu s'en rendre compte dans le cas d'estimations de lois $g(\delta')$ connues (cf § 4), des remontées anormales (certes moins fortes) se produisant en début et fin d'histogramme.

On a reporté sur le tableau 4 le gain moyen $\bar{\delta} = \Sigma b_i \delta_i / \Sigma b_i$, l'écart-type $\sigma^2 = \Sigma b_i (\delta_i - \bar{\delta})^2 / \Sigma b_i$, associé à chacun des 10 s.e.u. expérimentaux, ainsi que la somme des b_i qui est très voisine de 1 (comprise entre 0,99

et 1,006), les b_i étant calculés à partir de l'analyse factorielle soit du tableau k'_{IJ} , soit du tableau k_{IJ} . On voit que l'estimation $\bar{\delta}$ du gain moyen du premier étage, qui est un peu plus forte dans le second cas, et qui varie soit entre 7,29 et 7,61 (premier cas), soit entre 7,40 et 7,66 (second cas), est bien, quoiqu'un peu plus faible, de l'ordre de grandeur du gain 7,65 adopté pour la normalisation des s.e.u. expérimentaux (cf §§ 3.1, 3.2).

n° du s.e.u.	1	2	3	4	5	6	7	8	9	10	moyenne
1° cas Σb_i	1,000	0,999	0,996	1,003	1,002	1,005	1,006	0,996	1,005	0,993	1,0006
	7,61	7,56	7,53	7,47	7,58	7,55	7,39	7,42	7,29	7,41	7,48
	3,31	3,26	3,22	3,34	3,36	3,43	3,34	3,28	3,31	3,24	3,31
2° cas Σb_i	0,994	0,996	0,994	0,998	0,993	0,999	1,001	0,994	0,991	0,990	0,9949
	7,66	7,59	7,54	7,51	7,64	7,60	7,43	7,40	7,41	7,43	7,52
	3,26	3,23	3,20	3,33	3,30	3,40	3,34	3,27	3,21	3,21	3,275

Tableau n° 4 : Estimation du gain moyen $\bar{\delta}$ du premier étage et de l'écart-type associé σ .

1° cas : estimation faite à partir du tableau k'_{IJ}
 2° cas : estimation faite à partir du tableau k_{IJ}

Remarques

1) L'histogramme moyen permettant d'estimer $g(\delta')$ (qu'il soit obtenu à partir de k_{IJ} ou k'_{IJ}) et qui est calculé à partir des 7 premiers facteurs est pratiquement identique à l'histogramme moyen calculé à partir des 5 ou 6 premiers facteurs.

2) Au lieu d'introduire les facteurs dans l'ordre naturel (i.e. par valeurs propres décroissantes), on peut les introduire en fonction de leur importance (mesurée par la valeur absolue du cosinus entre le s.e.u. y_J considéré, et l'axe factoriel $u_{\alpha J}$), ce qui revient à classer les facteurs suivant les valeurs décroissantes de $|F_\alpha(y)|$, $F_\alpha(y)$ étant la valeur du $\alpha^{\text{ème}}$ facteur pour y_J (i.e. l'abscisse de la projection de y_J sur $u_{\alpha J}$).

Suivant les s.e.u. expérimentaux, on obtient au rang 5, soit l'ordre 2,1,3,4,5 ; soit l'ordre 2,3,1,4,5, sauf pour le premier s.e.u. pour lequel on a l'ordre 2,1,3,4,6, le cinquième facteur s'introduisant dans ce dernier cas immédiatement après. A partir du rang 6, on obtient soit le cinquième facteur (1° s.e.u.), soit le sixième, ou le septième, ou même le dixième (sixième s.e.u.), tandis qu'à partir du rang 7 s'introduisent des facteurs de rang élevé ; au rang $r=7$ (et même au rang $r=6$ pour le sixième facteur), la projection de chaque s.e.u. y_J sur l'espace associé aux r facteurs les plus liés à y_J est à l'extérieur du convexe \hat{C}_r correspondant. Au rang $r=5$ ($r=6$ pour le premier facteur) l'espace des r meilleurs facteurs coïncide avec l'espace des r premiers facteurs. Ces considérations expliquent d'une certaine façon pourquoi, dans l'estimation de $g(\delta')$, l'on obtient pratiquement le même histogramme moyen, que l'on raisonne avec 5,6 ou 7 facteurs (cf remarque 1).

3) Pour tester la validité des résultats obtenus, on a, outre l'estimation de lois $g(\delta')$ connues (cf § 4), simulé des s.e.u. différant des s.e.u. expérimentaux par une erreur traduisant les fluctuations d'échantillonnage. De façon précise, on a posé pour un s.e.u. expérimental y_J donné :

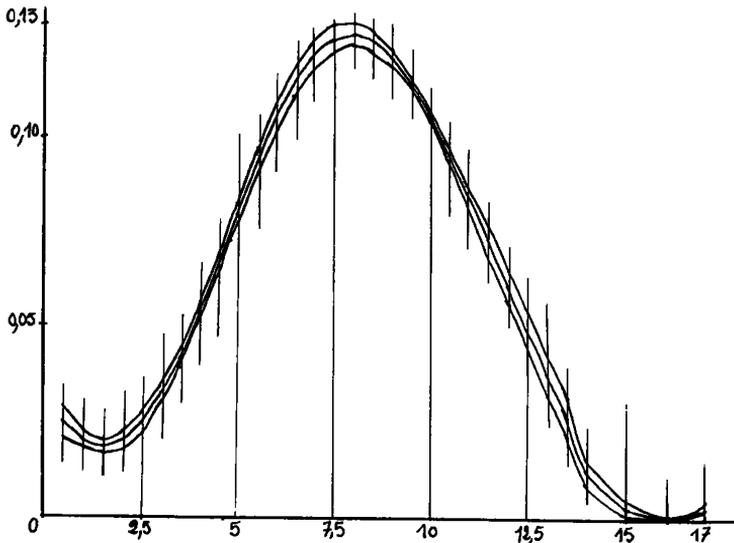


Figure 4 : Histogramme moyen $g'(\delta')$ des \bar{b}'_i associé aux 10 s.e.u. expérimentaux. On a encadré g' par les deux courbes g'' et g''' obtenues à partir de g' en retranchant et rajoutant respectivement à \bar{b}'_i l'écart type σ_i associé aux 10 s.e.u. expérimentaux. Pour tracer ces courbes, on a joint de façon continue les valeurs obtenues aux extrémités du découpage adopté, extrémités caractérisées par les traits verticaux sur le graphique.

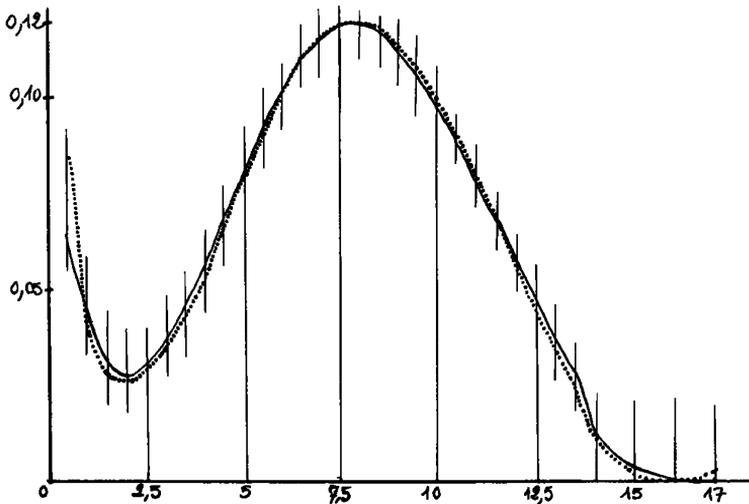


Figure 5 : Estimation de la loi $g(\delta')$ moyenne associée aux 10 s.e.u. expérimentaux.

..... : Estimation à partir du tableau k'_{IJ}

———— : Estimation à partir du tableau k_{IJ}

Pour tracer les courbes relatives à ces deux estimations, on a joint de façon continue les valeurs estimées aux extrémités du découpage adopté, extrémités caractérisées par les traits verticaux sur le graphique.

$$\forall j \in J : y'_j = y_j + \sqrt{y'_j} e_j$$

où le terme $\sqrt{y'_j} e_j$ traduit les erreurs dues aux fluctuations de comptage, à la sortie du P. M., e_j étant une erreur normale centrée de variance ϵ^2 . La dispersion des 10 s.e.u. expérimentaux correspondant à un écart-type ϵ égal à $2,1 \times 10^{-3}$, on a adopté pour ϵ des valeurs comprises entre 10^{-2} et 10^{-4} . Les estimations de la courbe $g(\delta')$ obtenues à partir des s.e.u. y'_j associés aux différentes valeurs de ϵ ainsi choisies ne diffèrent sensiblement pas de l'estimation obtenue à partir de y_j ($\epsilon = 0$).

4 Estimation de lois $g(\delta')$ connues

4.1 Introduction

Pour tester la méthode de régression développée précédemment, on se fixe ici une loi $g(\delta')$, par exemple une loi normale, de moyenne et de variance connues ; l'on en déduit (un découpage ayant été adopté, par exemple le découpage en 31 classes considéré aux §§ 3.3 et 3.4, découpage que nous conserverons) par la formule (5) les coefficients b_i et donc le s.e.u. associé à $g(\delta')$ qui est donné par la formule :

$$y_J = \Sigma \{ b_i k_{iJ} \mid i = 1, p \} \quad (17)$$

(en fait, si l'on fait l'analyse des correspondances du tableau k'_{IJ} associé aux s.e.u. théoriques ramenés chacun à avoir masse totale 1 (cf § 3.4) on remplacera dans la formule (17) k_{iJ} par k'_{iJ} ; c'est ce que l'on a fait ici).

On considérera également les s.e.u. "perturbés" y'_j déduits de y_j par la formule :

$$\forall j \in J : y'_j = y_j + \sqrt{y'_j} e_j \quad (18)$$

où les e_j sont des erreurs normales centrées indépendantes, et de même variance ϵ^2 , le terme en $\sqrt{y'_j}$ ayant été introduit de façon à se rapprocher des conditions expérimentales quand on enregistre un s.e.u., les erreurs de comptage étant de l'ordre de $\sqrt{y'_j}$.

L'on va essayer de reconstituer les b_i par la méthode de régression exposée aux §§ 2.2 et 2.3, et appliquée au § 3 dans le cas des 10 s.e.u. expérimentaux, les b_i étant estimés à partir de y'_j (dont y_j sera considéré comme un cas particulier, correspondant à une variance ϵ^2 nulle).

On peut d'abord noter que si l'on essaye d'estimer les coefficients b_i à partir de y_j , et si l'on garde tous les facteurs (y compris bien sûr le facteur trivial constant) issus de l'analyse des correspondances de k_{IJ} (ou de k'_{IJ}) (ce qui revient à prendre dans les formules (11), (13) à (15), $r+1 = \text{Card } I = p$) on retrouve exactement les coefficients b_i (sans avoir à imposer les contraintes (14)) puisque l'on s'est alors ramené à faire la régression usuelle, dans R_J muni de la métrique du chi-deux associée au tableau k_{IJ} (ou k'_{IJ}) analysé, de y_j par rapport aux $\{k_{iJ} \mid i = 1, p\}$.

Ce qui est intéressant, c'est de voir l'approximation réalisée quand on ne garde que 4, 5, 6 ou 7 facteurs, et de comparer cette approximation avec l'estimation des b_i obtenue à partir des s.e.u. y'_j perturbés,

cette estimation se faisant dans l'espace des r_0 premiers facteurs si y'_{J} se trouve dans le convexe C_{r_0} et à l'extérieur de C_{r_0+1} (cf § 2.2 *in fine*), r_0 dépendant bien sûr de ϵ ($r_0 = p-1$ si $\epsilon = 0$). On a ainsi un moyen de se rendre compte de la validité et des limites de la méthode de régression utilisée, i.e. de son aptitude à reconstituer à partir d'un s.e.u. y'_{J} perturbé la loi $g(\delta')$ sous-jacente.

4.2 Estimation de lois normales et de lois gamma

On a considéré des lois $g(\delta')$ de moyenne $\delta_0 = 7,55$ et d'écart-type $s_0 = 3,25$, ces deux valeurs étant proches des valeurs du gain moyen et de l'écart-type des lois $g(\delta')$ estimées à partir des 10 s.e.u. expérimentaux étudiés au § 3 (cf tableau 4) ; ce faisant les s.e.u. y'_{J} associés à ces lois $g(\delta')$ se projettent dans le plan 1-2 de l'analyse factorielle de k'_{IJ} au voisinage des 10 s.e.u. expérimentaux, i.e. au voisinage du second axe factoriel. On a étudié le cas d'une loi normale, et d'une loi gamma ($g(\delta') = a^t \delta'^{t-1} \exp(-a \delta') / \Gamma(t)$; a et t étant déterminés de telle sorte que $E(\delta') = t/a = \delta_0 = 3,25$; $\text{Var}(\delta') = t/a^2 = s_0^2 = (3,25)^2$).

Plusieurs valeurs de ϵ ont été considérées : $\epsilon = 0$; $\epsilon = 10^{-3}$; $\epsilon = 5 \times 10^{-3}$; $\epsilon = 10^{-2}$. Pour $\epsilon = 5 \times 10^{-3}$, on obtient une précision de la reconstitution du s.e.u. y'_{J} un peu moins bonne que pour les s.e.u. expérimentaux étudiés au § 3, tandis que pour $\epsilon = 10^{-3}$, elle est meilleure ; cette précision étant mesurée par le rapport $S^2 = \|y'_{\text{J}} - yy'_{\text{J}}\|^2 / \|y'_{\text{J}}\|^2$, (cf Méthodes de régression I § 1, Cahiers, Vol III n° 2, p. 150), yy'_{J} étant la reconstitution de y'_{J} , R_{J} étant muni de la métrique du χ^2 associée au tableau k'_{IJ} (ou k'_{IJ}) analysé (*).

Pour $\epsilon = 0$, la courbe $g(\delta')$ est pratiquement reconstituée avec quatre facteurs dans le cas de la loi normale, et avec cinq dans le cas de la loi gamma. Pour comparer les résultats obtenus avec ces diverses simulations, on a reporté sur le tableau 5 (tableau 5a dans le cas de la loi normale ; tableau 5b dans le cas de la loi gamma, où l'on a également retranscrit les résultats relatifs au cas où $\epsilon = 2,5 \times 10^{-3}$, cas non traité dans l'hypothèse de normalité) la valeur r_0 de r telle que y'_{J} soit à l'intérieur du convexe C_{r_0} et à l'extérieur de C_{r_0+1} (cf § 2.2), ainsi que les valeurs des statistiques $T^2 = \sum (b_i - \hat{b}_i)^2$, et $U^2 = \sum (b_i - \hat{b}_i)^2 / (b_i + \hat{b}_i)$ en fonction du nombre r ($r \leq r_0$) de facteurs considérés pour obtenir l'estimation \hat{b}_i de b_i .

On voit que dans le cas de la loi gamma, et pour les valeurs de ϵ non nulles, y'_{J} est à l'intérieur du convexe C_6 associé aux 6 premiers facteurs, les valeurs minimales de T^2 et U^2 étant obtenues pour $r = 5$ ou 6 facteurs. On a reporté sur la figure 6 la loi $g(\delta')$, et ses estimations, dans le cas où $\epsilon = 5 \times 10^{-3}$, à partir de $r = 5$ et 6 facteurs. On voit que l'on a une reconstitution convenable de $g(\delta')$.

En ce qui concerne la loi normale (cf fig. 7, où l'on a reporté $g(\delta')$ et ses estimations pour 4, 5 et 6 facteurs, toujours dans le cas où $\epsilon = 5 \times 10^{-3}$, la valeur de r_0 (pour $\epsilon \neq 0$) varie entre 6 et 8, tandis que les valeurs du nombre r de facteurs pour lequel T^2 et U^2 passent par un minimum varient

(*) S^2 est de l'ordre de $0,17 \times 10^{-2}$ pour $\epsilon = 5 \times 10^{-3}$, de $0,65 \times 10^{-4}$ pour $\epsilon = 10^{-3}$, alors que pour les dix spectres expérimentaux étudiés au § 3, il varie entre $0,16 \times 10^{-3}$ et $0,5 \times 10^{-3}$. En fait la dispersion des s.e.u. expérimentaux correspond, comme on l'a indiqué dans la remarque 3 du § 3.4, à une valeur de ϵ égale à $2,1 \times 10^{-3}$.

(toujours pour $\epsilon \neq 0$) entre 4 et 6. On peut noter que pour ϵ égal à 10^{-2} , r_0 vaut 7, tandis que pour ϵ égal à 5×10^{-3} , r_0 vaut 6, alors qu'*a priori*, on aurait pu s'attendre à ce que r_0 soit une fonction décroissante de ϵ . Ce résultat est probablement dû au générateur de nombres au hasard utilisé (*)

ϵ	0		10^{-2}		5×10^{-3}		10^{-3}	
	r_0		7		6		8	
r	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$
3	13,2	295	13,3	296	13,1	295	13,2	295
4	1,27	42	1,57	44	1,62	73	1,29	45
5	0,87	48	1,19	53	4,46	236	0,90	52
6	0,34	12	1,19	70	4,47	237	0,56	315
7	0,16	6,2	1,54	81	/	/	2,53	159
8	0,16	6,3	/	/	/	/	5,23	198
9	0,15	4,7	/	/	/	/	/	/

Tableau 5 : a) cas d'une loi normale

ϵ	0		10^{-2}		5×10^{-3}		$2,5 \times 10^{-3}$		10^{-3}	
	r_0		6		6		6		6	
r	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$	$10^4 T^2$	$10^4 U^2$
3	24,3	512	20,0	479	20,60	480	20,4	480	20,3	479
4	3,86	114	3,99	132	4,20	116	4,21	118	3,89	115
5	0,31	42	0,83	67	4,15	115	1,95	56	0,340	43,3
6	0,19	19	1,30	22	7,24	351	3,23	155	0,339	42,1
7	0,19	18	/	/	/	/	/	/	/	/
8	0,13	5,0	/	/	/	/	/	/	/	/
9	0,12	5,4	/	/	/	/	/	/	/	/

Tableau 5 : b) cas d'une loi gamma

Estimation de lois $g(\delta')$ connues (de moyenne 7.55 et écart-type 3.25), à partir de s.e.u. y'_j perturbés par une erreur de variance ϵ^2 .

Pour chaque valeur de ϵ , on a reporté la valeur r_0 telle que y'_j est à l'intérieur du convexe C_{r_0} et à l'extérieur de C_{r_0+1} , ainsi que les valeurs des statistiques T^2 et U^2 en fonction du nombre r de facteurs conservé.

(*) Ce générateur est le générateur en bibliothèque de l'UNIVAC 1110 de la faculté des sciences d'Orsay; il tire au hasard deux nombres x_1 et x_2 distribués uniformément sur $(0,1)$ et en déduit un nombre au hasard e gaussien centré de variance ϵ^2 par la formule classique $e = \epsilon(-2 \log x_1)^{1/2} \cos 2\pi x_2$. Les nombres x_1 et x_2 sont tirés à partir de la formule $x_{n+1} = (5 \times 10^{13})x_n \text{ Mod}(2^{35})$. On tire d'abord les 95 premiers nombres e_j associés à $\epsilon = 10^{-2}$, puis les 95 nombres suivants associés à $\epsilon = 5 \times 10^{-3}$ et enfin les 95 derniers nombres associés à $\epsilon = 10^{-3}$.

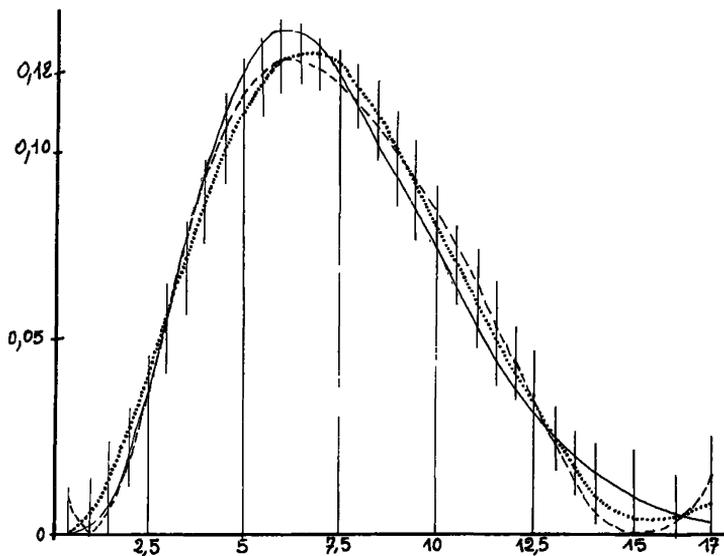


Figure 6 : Cas d'une loi gamma

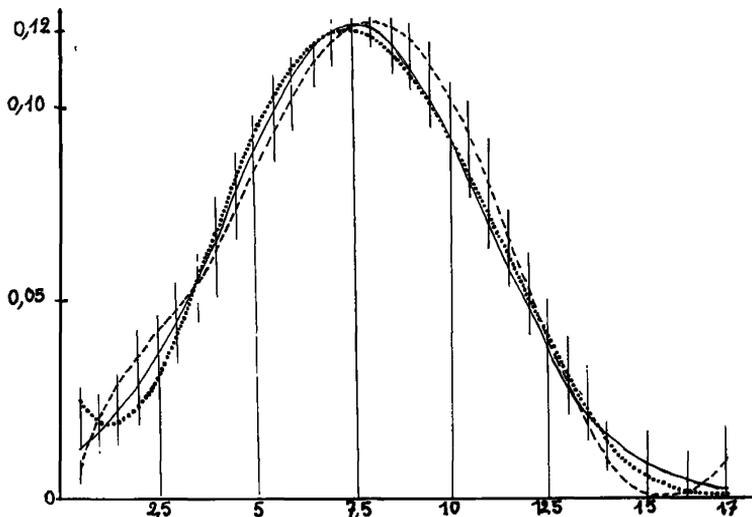


Figure 7 : Cas d'une loi normale

Figures 6 et 7 : Estimation de la loi gamma (Fig. 6) ou de la loi normale (Fig. 7) $g(\delta')$ de moyenne 7,55 et d'écart type 3,25 dans le cas d'un s.e.u. perturbé par une erreur normale d'écart type $\varepsilon = 5 \times 10^{-3}$. On a tracé la courbe $g(\delta')$ ainsi que ses estimations obtenues avec 4, 5 ou 6 facteurs.

Pour tracer ces courbes, on a joint de façon continue les valeurs obtenues aux extrémités du découpage adopté, extrémités caractérisées par les traits verticaux sur le graphique.

— loi $g(\delta')$
 estimation avec 4 ou 5 facteurs (Fig. 6), avec 4 facteurs (Fig. 7)
 ----- estimation avec 6 facteurs (Fig. 6), avec 5 ou 6 facteurs (Fig. 7)

Il est par ailleurs intéressant de noter que, comme dans le cas de la loi γ , T^2 et U^2 ne passent pas forcément par leur minimum pour la même valeur de r , et que même si c'est le cas, la valeur associée de r peut être différente (plus petite) que r_0 . Néanmoins les estimations obtenues en gardant r_0 facteurs semblent les meilleures pour la reconstitution globale de la courbe $g(\delta')$ testée sauf aux extrémités, où l'on a une légère remontée anormale de l'estimation de $g(\delta')$ (cf fig. 6 et 7).

En conclusion, l'on obtient avec la méthode de régression utilisée une bonne reconstitution des lois normale et γ testées.

4.3 Remarques

1) Si l'on se place dans l'espace des $r = r_0 + 1$ premiers facteurs, la régression usuelle (i.e. sans imposer les contraintes de positivité (14)) qui pour $r \leq r_0$ donnait des résultats à peu près convenables (courbes unimodales avec quelques valeurs négatives aux extrémités, faibles en valeur absolue) donne des résultats complètement aberrants : courbe ayant 5 ou 6 *extrema*, dont certains négatifs et de forte amplitude. En imposant les contraintes (14), ce qui protège dans une certaine mesure la régression, on obtient une estimation de $g(\delta')$ unimodale, moins régulière que pour $r = r_0$, avec, quand il y en a, des remontées plus importantes aux extrémités ; les valeurs des statistiques T^2 et U^2 sont alors plus élevées que pour r_0 .

2) On peut noter sur les figures 6 et 7 de légères remontées, anormales, au début et à la fin des histogrammes estimés, l'estimation d'une loi $g(\delta')$ est donc peu précise pour les plus faibles, ou les plus fortes valeurs de δ' , comme on l'a déjà signalé au § 3.4 dans le cas des 10 s.e.u. expérimentaux.

3) Dans toutes les régressions effectuées aux §§ 3 et 4 (s.e.u. expérimentaux, s.e.u. perturbés à partir de lois $g(\delta')$ connues) les estimations de $g(\delta')$ avec deux ou trois facteurs sont unimodales avec un maximum très plat ; de plus, ces estimations sont relativement voisines ; ce résultat se comprend aisément dans le cas de deux facteurs, puisque d'une part l'estimation obtenue dépend essentiellement des deux premiers moments des lois $g(\delta')$ considérées, moments qui sont du même ordre de grandeur dans tous les essais effectués, et que d'autre part, on a "centré" la parabole du plan 1-2 sur la moyenne (cf § 3.3).

4) Si l'on essaie de reconstituer une distribution $g(\delta')$ bimodale, mélange de deux lois normales, la méthode de régression précédente permet, malgré un certain lissage, de retrouver la bimodalité si ces deux lois sont bien séparées. Si ces deux lois sont imbriquées, on n'a retrouvé (dans les essais effectués) la bimodalité, quand ϵ est nul, qu'à partir du 13^e facteur. Comme pour des valeurs de ϵ non nulles, comprises entre 10^{-2} et 10^{-4} , la valeur maximale r_0 pour laquelle les s.e.u. perturbés associés y_j sont à l'intérieur du convexe C_{r_0} , varie entre 6 et 9 suivant la valeur de ϵ , l'estimation obtenue pour $g(\delta')$ est unimodale. Pour retrouver la bimodalité, et avoir une estimation convenable de $g(\delta')$, il faut, comme on l'a vérifié, soit considérer des s.e.u. perturbés y_j par des erreurs de variance plus faible ($\epsilon = 10^{-5}$), soit considérer des s.e.u. explicatifs de variance plus faible.

5) On a également, dans le cas d'une loi $g(\delta')$ normale, ou mélange de deux lois normales (cf remarque précédente), essayé de reconstituer cette loi en se plaçant dans l'espace des premiers facteurs de l'analyse en composantes principales du tableau k_{IJ} associé à $g(\delta')$ (cf § 2.3, remarque 2). On a obtenu des estimations de $g(\delta')$ pratiquement identiques à celles obtenues à partir de l'analyse des correspondances, mais avec des ajustements du même ordre de grandeur, ou un peu moins bons.

BIBLIOGRAPHIE

- [1] BALLINI, J.P. (1974) : Spectres d'impulsions à "un photoélectron" des photomultiplicateurs, *Nuclear Instruments And Methods*, 116, pp 109-121.
- [2] BALLINI, J.P., CAZES, P., TURPIN, P.Y. (1976) :
Single-electron multiplication statistics as a combination of Poissonian pulse height distributions using constraint regression methods, *Nuclear Instruments And Methods*, 134, pp 319-330.
- [3] LOMBARD, F.J., MARTIN, F. (1961) :
Statistics of electron multiplication, *The Review of Scientific Instruments*, Vol 32, n° 2, pp 200-201.
- [4] PRESCOTT, J.R. (1966) : *A Statistical model for photomultiplier single-electron statistics*, *Nuclear Instruments And Methods*, 39, p 173.