

REVUE DE STATISTIQUE APPLIQUÉE

A. ROSENGARD

Le groupage des points dans l'estimation des paramètres inconnus d'une régression

Revue de statistique appliquée, tome 13, n° 1 (1965), p. 39-47

http://www.numdam.org/item?id=RSA_1965__13_1_39_0

© Société française de statistique, 1965, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

LE GROUPEMENT DES POINTS DANS L'ESTIMATION DES PARAMÈTRES INCONNUS D'UNE RÉGRESSION ⁽¹⁾

A. ROSENGARD
Ingénieur au C.E.A. (Saclay)

Dans cet article nous énonçons un théorème peu connu qui permet, en groupant en général les expériences en autant de points qu'il y a de paramètres à estimer, de trouver un plan d'expérience simple permettant de calculer rapidement les paramètres inconnus, et de rendre maximum la précision sur un des paramètres, quand l'ordre de grandeur de ceux-ci est connu.

I - INTRODUCTION

Il arrive souvent, notamment en physique, qu'on ait à estimer les paramètres a , b , c , d'une fonction d'une variable x :

$$y = f(x, a, b, c) \quad (1)$$

Nous prenons ici 3 paramètres afin de fixer les idées ; de façon générale, on a k paramètres inconnus.

Nous supposons qu'on peut fixer, sans erreur appréciable, la valeur de x ; y est en revanche mesurée avec une certaine erreur ε aléatoire ($E(\varepsilon) = 0$). On pourra admettre que les erreurs ε sont distribuées suivant une loi normale de moyenne nulle et d'écart-type constant σ (on suppose, afin d'alléger l'exposé que σ est indépendant de x). Tout ce qui suit pourra s'adapter au cas où σ dépend de x de façon connue.

Notre modèle peut donc s'écrire :

$$y = f(x, a, b, c) + \varepsilon \quad (2)$$

Il est bien entendu que tout ce qui suivra n'est valable que si la forme analytique de $f(x, a, b, c)$ est *RIGOREUSEMENT CONNUE*.

Nous supposons que l'on peut reproduire autant de fois qu'on veut les mesures de y pour une valeur de x donnée, et que le domaine de variation de x est fermé et borné (ce qui est toujours le cas en pratique).

Nous désirons avoir, pour un nombre donné n d'expériences, l'estimation la plus précise sur *un des paramètres*, a , par exemple. Nous

(1) Il s'agit de régression au sens général du mot, et non spécialement de régression linéaire.

voulons, pour celà, savoir comment répartir les positions de x où nous ferons les mesures de y . Or, il existe un résultat fondamental, donné par Hermann Chernoff (réf. 1), qui va nous permettre de donner en général le plan d'estimation optimum. Ce plan va de plus nous permettre de donner les estimations de a , b , c , avec leur précision, et cela par des calculs relativement aisés.

Ce résultat est le suivant :

Un plan localement optimum sera obtenu en faisant les expériences en au plus k points (s'il y a k paramètres inconnus) x_1, \dots, x_k et en les répétant dans certaines proportions n_1, \dots, n_k avec $n_1 + \dots + n_k = n$, et ceci à condition que le domaine Δ où l'on peut faire varier x soit fermé et borné.

La démonstration de ce résultat est assez délicate, et il ne nous est même pas possible, de l'esquisser dans le cadre de cet article. Nous renvoyons les lecteurs intéressés à (Réf. 1).

Il faut remarquer qu'il s'agit d'un plan localement optimum : autrement dit ce plan ne sera optimum que si a , b , c sont dans le voisinage de trois valeurs a_0 , b_0 , c_0 qui sont les ordres de grandeurs présumées de a , b , c . Si on ne connaît pas déjà ces ordres de grandeurs, il conviendra de faire une première expérience en se donnant x_1 , x_2 et x_3 , avec n faible, et d'appliquer la méthode d'estimation que nous allons exposer. Une fois connus les ordres de grandeur, on pourra donner le plan optimum.

II - EXEMPLE

Avant d'exposer notre méthode, nous allons expliquer le théorème dans un cas particulier, celui de la régression linéaire (voir aussi réf. 1).

Soit à estimer les paramètres a , b , de la régression linéaire

$$y = a x + b + \varepsilon \quad (3)$$

où ε est l'erreur, supposée normale et d'écart-type constant.

Supposons que le domaine de variation de x soit $-1 \leq x \leq +1$. Si on veut estimer a de façon optimum, il est bien connu qu'on a intérêt à faire $\frac{n}{2}$ expériences en $x = +1$ et $\frac{n}{2}$ en $x = -1$: nous sommes dans le cas général où il faut faire autant de groupages qu'il y a de paramètres inconnus ($k = 2$).

Cependant, si nous voulions estimer b , a ne nous intéressant pas, il est évident qu'il faudrait faire n expériences en $x = 0$ (a ne serait pas connu). Nous sommes alors dans un cas particulier où il ne faut faire qu'un seul groupage ($1 \leq 2$).

Dans l'exposé qui suivra, nous nous placerons toujours dans le cas général, où il faut faire autant de groupages qu'il y a de paramètres inconnus.

Les conditions portant sur le domaine de variation de x sont ici évidentes. Si, par exemple, nous choisissons le domaine ouvert $-1 < x < +1$, à tout plan consistant à faire des expériences en $-1 + \varepsilon$ et $1 - \varepsilon$, on peut faire correspondre le plan donnant une meilleure précision : $1 + \frac{\varepsilon}{2}$

et $1 - \frac{\varepsilon}{2}$. De même si le domaine n'est pas borné, à tout plan consistant à faire les expériences en $-A, +A$, on peut faire correspondre le plan $-2A, +2A$ dont la précision est meilleure.

Enfin notons que dans notre exemple le plan est uniformément optimum, car il ne dépend pas de a et b : ce n'est pas bien entendu le cas général.

III - EXPOSE DE LA METHODE DE GROUPAGE

III.1 - Estimation des paramètres

Nous savons que le plan optimum va consister à faire des expériences en $x_1, x_2, x_3 \in \Delta$, avec en chaque position respectivement n_1, n_2, n_3 pointés, $n_1 + n_2 + n_3 = n$. Dans ce cas l'estimation de a, b, c est immédiate :

Soit y_i^j le résultat de la $j^{\text{ème}}$ expérience relative à la position i ($i = 1, 2, 3$; $j = 1, 2, \dots, n_i$).

On peut calculer les moyennes de ces trois groupes de points :

$$y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_i^j, \quad (i = 1, 2, 3) \quad (4)$$

alors a, b, c seront calculés en résolvant le système :

$$\begin{cases} y_1 = f(x_1, a, b, c) \\ y_2 = f(x_2, a, b, c) \\ y_3 = f(x_3, a, b, c) \end{cases} \quad (5)$$

S'il y a plusieurs solutions, seule est bien entendu à prendre en compte celle qui se trouve dans le voisinage de a_0, b_0, c_0 .

III.2 - Recherche de la précision asymptotique du plan :

Soit G une grandeur. Nous désignons par δG l'erreur (aléatoire) faite sur G . Supposons que $\delta a, \delta b, \delta c$ soient faibles, ce qui est vrai si n est assez grand (du reste on veut pratiquement qu'il en soit ainsi). Alors on peut écrire : (étant entendu que $a \simeq a_0, b \simeq b_0, c \simeq c_0$).

$$\delta f(x, a, b, c) = \frac{\partial f}{\partial a} \delta a + \frac{\partial f}{\partial b} \delta b + \frac{\partial f}{\partial c} \delta c \quad (6)$$

Si on applique (6) au système (5), on voit que :

$$\begin{aligned} \delta y_1 &= \frac{\partial f_1}{\partial a} \delta a + \frac{\partial f_1}{\partial b} \delta b + \frac{\partial f_1}{\partial c} \delta c \\ \delta y_2 &= \frac{\partial f_2}{\partial a} \delta a + \frac{\partial f_2}{\partial b} \delta b + \frac{\partial f_2}{\partial c} \delta c \end{aligned} \quad (7)$$

$$\delta y_3 = \frac{\partial f_3}{\partial a} \delta a + \frac{\partial f_3}{\partial b} \delta b + \frac{\partial f_3}{\partial c} \delta c$$

(f_i désigne ici $f(x_i, a, b, c)$)

La solution de ce système peut se présenter ainsi :

$$\delta a = \frac{1}{D} \begin{vmatrix} \delta y_1 & \frac{\partial f_1}{\partial b} & \frac{\partial f_1}{\partial c} \\ \delta y_2 & \frac{\partial f_2}{\partial b} & \frac{\partial f_2}{\partial c} \\ \delta y_3 & \frac{\partial f_3}{\partial b} & \frac{\partial f_3}{\partial c} \end{vmatrix}$$

$$\delta b = \frac{1}{D} \begin{vmatrix} \frac{\partial f_1}{\partial a} & \delta y_1 & \frac{\partial f_1}{\partial c} \\ \frac{\partial f_2}{\partial a} & \delta y_2 & \frac{\partial f_2}{\partial c} \\ \frac{\partial f_3}{\partial a} & \delta y_3 & \frac{\partial f_3}{\partial c} \end{vmatrix} \quad (8)$$

$$\delta c = \frac{1}{D} \begin{vmatrix} \frac{\partial f_1}{\partial a} & \frac{\partial f_1}{\partial b} & \delta y_1 \\ \frac{\partial f_2}{\partial a} & \frac{\partial f_2}{\partial b} & \delta y_2 \\ \frac{\partial f_3}{\partial a} & \frac{\partial f_3}{\partial b} & \delta y_3 \end{vmatrix}$$

où :

$$D = \begin{vmatrix} \frac{\partial f_1}{\partial a} & \frac{\partial f_1}{\partial b} & \frac{\partial f_1}{\partial c} \\ \frac{\partial f_2}{\partial a} & \frac{\partial f_2}{\partial b} & \frac{\partial f_2}{\partial c} \\ \frac{\partial f_3}{\partial a} & \frac{\partial f_3}{\partial b} & \frac{\partial f_3}{\partial c} \end{vmatrix} \quad (9)$$

Afin d'alléger la présentation, nous écrivion (8) sous la forme :

$$\left\{ \begin{array}{l} \delta a = A_1^1 \delta y_1 + A_1^2 \delta y_2 + A_1^3 \delta y_3 \\ \delta b = A_2^1 \delta y_1 + A_2^2 \delta y_2 + A_2^3 \delta y_3 \\ \delta c = A_3^1 \delta y_1 + A_3^2 \delta y_2 + A_3^3 \delta y_3 \end{array} \right. \quad (10)$$

On voit, sous cette forme, que comme les erreurs δy_i sont normales $(0, \frac{\sigma}{\sqrt{n_i}})$, δa , δb , δc sont normales et de variance respective σ_a^2 , σ_b^2 , σ_c^2 :

$$\begin{aligned} \sigma_a^2 &= \sigma^2 \left[\frac{(A_1^1)^2}{n_1} + \frac{(A_1^2)^2}{n_2} + \frac{(A_1^3)^2}{n_3} \right] \\ \sigma_b^2 &= \sigma^2 \left[\frac{(A_2^1)^2}{n_1} + \frac{(A_2^2)^2}{n_2} + \frac{(A_2^3)^2}{n_3} \right] \\ \sigma_c^2 &= \sigma^2 \left[\frac{(A_3^1)^2}{n_1} + \frac{(A_3^2)^2}{n_2} + \frac{(A_3^3)^2}{n_3} \right] \end{aligned} \quad (11)$$

III.3 - Optimisation du Plan

Il nous faut trouver x_1 , x_2 , x_3 , n_1 , n_2 , n_3 ($n_1 + n_2 + n_3 = n$) tels que σ_a^2 soit minimum ; pour cela nous allons procéder en deux étapes, en fixant d'abord x_1 , x_2 , x_3 .

Il faut donc minimiser :

$$z = \frac{(A_1^1)^2}{n_1} + \frac{(A_1^2)^2}{n_2} + \frac{(A_1^3)^2}{n_3} \quad (12)$$

avec $n_1 + n_2 + n_3 = n$.

Ce minimum s'obtient immédiatement par la méthode des multipliateurs de Lagrange :

$$\begin{aligned} n_1 &= n \frac{|A_1^1|}{|A_1^1| + |A_1^2| + |A_1^3|} \\ n_2 &= n \frac{|A_1^2|}{|A_1^1| + |A_1^2| + |A_1^3|} \\ n_3 &= n \frac{|A_1^3|}{|A_1^1| + |A_1^2| + |A_1^3|} \end{aligned} \quad (13)$$

Le minimum de z est alors :

$$z_m = \frac{1}{n} (|A_1^1| + |A_1^2| + |A_1^3|)^2 \quad (14)$$

donc :

$$(\sigma_1)_m = \frac{\sigma}{\sqrt{n}} \left[|A_1^1| + |A_1^2| + |A_1^3| \right]$$

Cette quantité, calculée avec a_o, b_o, c_o dépend de $x_1, x_2, x_3 \in \Delta$: il faut chercher x_1, x_2, x_3 la rendant minimum. Cette dernière partie est souvent délicate au point de vue numérique.

III.4 - Remarque concernant l'utilisation d'un ordinateur

Si on désire se servir de la programmation pour effectuer les calculs numériques, on n'a pas intérêt à utiliser le système (8) pour avoir les quantités A_l^k ; en observant que $A_1^1 = \frac{\partial a}{\partial y_1} = \frac{\delta a}{\delta y_1}$ etc.. On voit que ces 9 quantités sont les solutions des 3 systèmes linéaires suivants :

$$\left\{ \begin{array}{l} \frac{\partial f_1}{\partial a} A_1^1 + \frac{\partial f_1}{\partial b} A_1^2 + \frac{\partial f_1}{\partial c} A_1^3 = 1 \\ \frac{\partial f_2}{\partial a} A_1^1 + \frac{\partial f_2}{\partial b} A_1^2 + \frac{\partial f_2}{\partial c} A_1^3 = 0 \\ \frac{\partial f_3}{\partial a} A_1^1 + \frac{\partial f_3}{\partial b} A_1^2 + \frac{\partial f_3}{\partial c} A_1^3 = 0 \end{array} \right. \quad (15)$$

$$\left\{ \begin{array}{l} \frac{\partial f_1}{\partial a} A_2^1 + \frac{\partial f_1}{\partial b} A_2^2 + \frac{\partial f_1}{\partial c} A_2^3 = 0 \\ \frac{\partial f_2}{\partial a} A_2^1 + \frac{\partial f_2}{\partial b} A_2^2 + \frac{\partial f_2}{\partial c} A_2^3 = 0 \\ \frac{\partial f_3}{\partial a} A_2^1 + \frac{\partial f_3}{\partial b} A_2^2 + \frac{\partial f_3}{\partial c} A_2^3 = 0 \end{array} \right. \quad (16)$$

$$\left\{ \begin{array}{l} \frac{\partial f_1}{\partial a} A_3^1 + \frac{\partial f_1}{\partial b} A_3^2 + \frac{\partial f_1}{\partial c} A_3^3 = 0 \\ \frac{\partial f_2}{\partial a} A_3^1 + \frac{\partial f_2}{\partial b} A_3^2 + \frac{\partial f_2}{\partial c} A_3^3 = 0 \\ \frac{\partial f_3}{\partial a} A_3^1 + \frac{\partial f_3}{\partial b} A_3^2 + \frac{\partial f_3}{\partial c} A_3^3 = 0 \end{array} \right. \quad (17)$$

IV - APPLICATION DE LA METHODE PRECEDENTE A UN EXEMPLE

Il arrive souvent en physique qu'un phénomène puisse être représenté de façon très précise dans un intervalle Δ donné par une équation de la forme⁽¹⁾ :

$$\Delta : 0 \leq x \leq \lambda \quad y = ax(1 + cx) + \varepsilon = ax + bx^2 + \varepsilon$$

ε est l'erreur, c est en général petit : ceci correspond à une correction $1 + cx$ sur le coefficient a .

Nous allons chercher le plan *optimum d'estimation de a*

On a :

$$\frac{\partial y}{\partial a} = x, \quad \frac{\partial y}{\partial b} = x^2, \quad D = \begin{vmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \end{vmatrix} = x_1 x_2 (x_2 - x_1)$$

En appliquant (8), il vient :

$$a = \frac{1}{D} \begin{vmatrix} \delta y_1 & x_1^2 \\ \delta y_2 & x_2^2 \end{vmatrix} \quad \text{donc} \quad \sigma_a^2 = \frac{\sigma^2}{D^2} \left[\frac{x_2^4}{n_1} + \frac{x_1^4}{n_2} \right]$$

$$b = \frac{1}{D} \begin{vmatrix} x_1 & \delta y_1 \\ x_2 & \delta y_2 \end{vmatrix} \quad \text{donc} \quad \sigma_b^2 = \frac{\sigma^2}{D^2} \left[\frac{x_2^2}{n_1} + \frac{x_1^2}{n_2} \right]$$

$$\text{Le minimum de } \sigma_a^2 \text{ a lieu pour } \begin{cases} n_1 = \frac{x_2^2}{x_1^2 + x_2^2} n \\ n_2 = \frac{x_1^2}{x_1^2 + x_2^2} n \end{cases}$$

et a pour valeur

$$(\sigma_a)_m = \frac{\sigma}{\sqrt{n}} \frac{x_1^2 + x_2^2}{x_1 x_2 |x_2 - x_1|}$$

Il reste à trouver le minimum de φ

$$\varphi = \frac{x_1^2 + x_2^2}{x_1 x_2 (x_2 - x_1)} \quad \text{où } 0 \leq x_1 \leq x_2 \leq \lambda$$

(nous supposons que $x_2 \geq x_1$, ce qui est toujours possible).

 (1) Remarquons que le modèle peut prendre la forme d'un modèle linéaire si l'on écrit : $u = y/x = a + bx + \varepsilon/x$, mais cette fois l'écart-type de l'erreur n'est plus constant.

On a :

$$\varphi = \frac{1}{x_1} - \frac{1}{x_2} + \frac{2}{x_2 - x_1}$$

Fixons x_2 ; pour $x_1 = 0$ et $x_1 = x_2$, $\varphi = +\infty$. Il y a donc au moins un minimum, et en ce minimum :

$$\frac{\partial \varphi}{\partial x_2} = -\frac{1}{x_1^2} + \frac{2}{(x_2 - x_1)^2} = 0 \quad \text{soit } x_1 = x_2 (\sqrt{2} - 1)$$

$$\text{et } \varphi = \frac{2(\sqrt{2} + 1)}{x_2}, \text{ qui est minimum pour } x_2 = \lambda.$$

En résumé, il faut prendre le plan suivant :

$$x_1 = \lambda (\sqrt{2} - 1) = 0,414 \lambda \quad n_1 = \frac{2 + \sqrt{2}}{4} n \simeq 0,85 n$$

$$x_2 = \lambda \quad n_2 = \frac{2 - \sqrt{2}}{4} n \simeq 0,15 n$$

La variable n_i étant entière, on arrondit à l'entier, le plus proche ; on a alors :

$$\left\{ \begin{array}{l} \sigma_a = \frac{4,8}{\lambda \sqrt{n}} \sigma \\ \sigma_b = \frac{6,3}{\lambda \sqrt{n}} \sigma \end{array} \right.$$

Si on remarque, en général, $a \gg b$, on voit que l'erreur relative sur b est bien supérieure à l'erreur relative sur a .

V - CONCLUSION

La méthode de groupage permet donc de trouver le plan optimum relativement à la précision d'un paramètre. De plus nous avons vu que l'estimation des paramètres revenait alors à la résolution d'un système d'équations immédiatement écrit, (alors que dans le cas général, on est d'abord obligé d'approcher la régression étudiée par un polynôme). Mais il est fondamental pour le succès de cette méthode d'être suffisamment sûr de l'exactitude de la forme analytique de la régression, sinon on est exposé aux pires déboires. Nous croyons bon d'insister sur ce point qui nous semble fondamental. S'il en est ainsi, dans le cas de nombreuses expériences où seuls les paramètres risquent de fluctuer, le plan conduit à de sérieuses économies.

Signalons enfin que si le coût de l'expérience dépend de x , on pourra traiter le problème en se donnant au lieu du nombre total d'expériences n , la somme que l'on veut allouer, et on cherchera à rendre l'estimation optimum avec cette nouvelle contrainte.

BIBLIOGRAPHIE

Nous avons rassemblé une bibliographie portant sur les plans d'estimation.

- [1] H. CHERNOFF - Locally optimal designs for estimating parameters .
Ann. Math. Stat. *24* (1953), p. 586/602.
- [2] G. ELFVING. - Optimum allocation in linear regression theory. Ann.
Math. Stat. *23* (1952), p. 255/262.
- [3] A. DE LA GARZA - Spacing of information in polynomial regression.
Ann. Math. Stat. *25* (1954), p. 123/130.
- [4] P.G. GUEST - The spacing of observations in polynomial regression .
Ann. Math. Stat. *29* (1958) p. 294/299.
- [5] J. KIEFER et J. WOLFOWITZ - Optimum designs in regression
problems. Ann. Math. Stat. *30* (1959) p. 271/294.
- [6] SOKOLOV - Theory of probability and its applications (Revue Russe
traduite en anglais). (1962), p. 89.