REVUE DE STATISTIQUE APPLIQUÉE

G. MATHERON

Théorie des droites de correspondance

Revue de statistique appliquée, tome 11, nº 4 (1963), p. 13-23
<http://www.numdam.org/item?id=RSA_1963__11_4_13_0>

THÉORIE DES DROITES DE CORRESPONDANCE

G. MATHERON
Ingénieur au Corps des Mines, détaché auprés du B.R.G.M.

INTRODUCTION

Dans les gisements d'Uranium, il est aisé et peu onéreux de mesurer les radioactivités de toutes les berlines de minerai sortant au jour. Ces radioactivités x sont en corrélation positive, généralement très forte, avec les teneurs y des berlines. Si les distributions sont normales ou lognormales, la droite de régression de y en x, ou droite de correspondance, permet donc, avec une précision d'autant meilleure que cette corrélation est plus forte, d'estimer la teneur y inconnue d'une berline donnée à partir de sa radioactivité mesurée x. Mais, en raison du prix de revient élevé du quartage, on ne disposera, en général, pour étalonner la droite de correspondance, que d'un nombre n relativement peu élevé de couples de valeurs expérimentales $(x_i \; y_i)$ en correspondance. Au contraire, puisque les radioactivités de toutes les berlines sont mesurées, la population des x est très bien connue, et les paramètres de leur distribution peuvent être considérés comme donnés. Le problème posé consiste à trouver la meilleure estimation de la teneur moyenne, et la précision de cette estimation, à partir des n couples expérimentaux $(x_i \; y_i)$ et des paramètres de. la loi des x connus à priori.

Nous nous limitons au cas où les n berlines étalonnées peuvent être considérées comme ayant été choisies au hasard parmi l'ensemble des berlines sortantes, de sorte que les x_i et les y_j sont indépendants les uns des autres, sauf dans le cas i = j. Nous allons traiter le problème par la méthode du maximum de vraisemblance dans le cas d'une distribution normale, puis dans le cas d'une distribution lognormale.

I - NOTATIONS DANS LE CAS GAUSSIEN

Soient x et y deux variables normales. Les paramètres de leur loi de distribution seront désignés par des lettres grecques :

$$
\left\{
\begin{aligned}
\mu_x &= E(x) \\
\mu_y &= E(y) \\
\sigma_x^2 &= D^2(x) \\
\sigma_y^2 &= D^2(y) \\
\sigma_{xy} &= E\,[(x - \mu_x)(y - \mu_y)] \\
\rho &= \frac{\sigma_{xy}}{\sigma_x \, \sigma_y} \\
\beta &= \rho \frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} \\
\sigma^2 &= \sigma_y^2 (1 - \rho^2) = \sigma_y^2 - \beta^2 \sigma_x^2
\end{aligned}
\right. \tag{1}
$$

On prendra garde que σ^2 désigne la variance de y lié par x, et β le coefficient de régression de y en x. L'équation théorique de la droite de correspondance (de régression) s'écrit :

$$E(y \mid x) = \mu_y + \beta (x - \mu_x) \tag{2}$$

Pour estimer les paramètres μ_y et β de cette droite, et la variance liée σ^2, nous disposons

1/ des valeurs exactes de μ_x et σ_x^2

2/ de n couples expérimentaux (x_i, y_i).

Nous désignerons par m_y, b et s^2 les estimateurs du maximum de vraisemblance de μ_y, β et σ^2. De même, certaines expressions formées à l'aide des valeurs expérimentales x_i y_i seront désignées par les lettres latines évoquant le paramètre (1) concerné. Ainsi, nous poserons :

$$
\left\{
\begin{aligned}
\bar{x} &= \frac{1}{n} \sum x_i \\
\bar{y} &= \frac{1}{n} \sum y_i \\
S_x^2 &= \frac{1}{n} \sum (x_i - \mu_x)^2 \\
s_x^2 &= \frac{1}{n} \sum (x_i - \bar{x})^2 \\
s_{xy} &= \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum (x_i - \mu_x)(y_i - \bar{y})
\end{aligned}
\right. \tag{3}
$$

Avec ces notations, il est clair que l'on a :

$$S_x^2 = s_x^2 + (\bar{x} - \mu_x)^2 \tag{4}$$

II - FORMATION DES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE (CAS NORMAL)

Le logarithme de la densité de fréquence de (x y) se met sous la forme :

$$\log f = -\log 2\pi - \frac{1}{2}\log \sigma_x^2 - \frac{1}{2}\log \sigma^2 - \frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{1}{2\sigma^2}[(y-\mu_y) - \beta(x-\mu_x)]^2$$

On obtient les estimateurs m_y, s^2 et b en écrivant que la fonction

$$\Phi(m_y, s^2, b) = n\log s^2 + \frac{1}{s^2}\sum_i [(y_i - m_y) - b(x_i - \mu_x)]^2 \qquad (5)$$

est minimale, soit

$$\frac{\partial\Phi}{\partial m_y} = \frac{\partial\Phi}{\partial s^2} = \frac{\partial\Phi}{\partial b} = 0$$

En tenant compte des notations (3), les trois équations ci-dessus se mettent sous la forme :

$$\left\{\begin{array}{l} m_y = \bar{y} - b(\bar{x} - \mu_x) \\ s^2 = \frac{1}{n}\sum [y_i - m_y - b(x_i - \mu x)]^2 \\ b = \frac{n\sum (x_i - \mu_x)(y_i - m_y)}{S_x^2} \end{array}\right.$$

Portons l'expression de m_y donnée par la première équation dans les deuxièmes membres des deux dernières. On obtient immédiatement :

$$\left\{\begin{array}{l} s^2 = \frac{1}{n}\sum [y_i - \bar{y} - b(x_i - \bar{x})]^2 \\ b = \frac{s_{xy}}{S_x^2} + b\frac{(\bar{x} - \mu x)^2}{S_x^2} \end{array}\right.$$

Compte tenu de (4), la dernière équation donne :

$$b = \frac{s_{xy}}{s_x^2}$$

D'où les estimateurs cherchés :

$$\left\{\begin{array}{l} m_y = \bar{y} - b(\bar{x} - \mu_x) \\ b = \frac{s_{xy}}{s_x^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ s^2 = \frac{1}{n}\sum_i [y_i - \bar{y} - b(x_i - \bar{x})]^2 \end{array}\right. \qquad (6)$$

On remarque que les estimateurs b et s^2 sont identiques à ceux que l'on obtient, classiquement, dans le cas où les paramètres μ_x et σ_x^2 ne sont pas connus à priori. Par contre, m_y peut différer substantiellement de la moyenne $\bar{y}$ des y_i. Pour un coefficient b positif, par exemple, m_y est pris inférieur à $\bar{y}$, comme il est naturel, dans le cas où les n échantillons x_i ont une moyenne $\bar{x}$ plus grande que la valeur vraie μ_x, et inversement.

III - CRITIQUE DES ESTIMATEURS OBTENUS

Examinons la valeur probable et la variance des estimateurs (6) ainsi obtenus. Comme b et s^2 ne diffèrent pas des estimateurs que l'on obtient dans le cas où μ_x et σ_x^2 ne sont pas connus à priori, nous bénéficions des résultats classiques(1) suivants :

1/ b et s^2 sont stochastiquement indépendants, et chacun d'eux est indépendant de $\bar{x}$ et de $\bar{y}$.

2/ La variable $\sqrt{n-1}\,\frac{\sigma_x}{\sigma}(b-\beta)$ est une variable de Student à n - 1 degrés de liberté.

3/ La variable $n\frac{s^2}{\sigma^2}$ est une variable χ^2 à n - 2 degrés de liberté.

De la propriété 2, nous déduisons immédiatement :

$$\left\{\begin{aligned} &E(b) = \beta \\ &D^2(b) = \frac{\sigma^2}{(n-3)\sigma_x^2} = \frac{\sigma_y^2(1-\rho^2)}{(n-3)\sigma_x^2} \end{aligned}\right. \qquad (7)$$

Ainsi, b est un estimateur sans biais. De la propriété 3, nous déduisons de la même manière :

$$\left\{\begin{aligned} &E(s^2) = \frac{n-2}{n}\sigma^2 = \frac{n-2}{n}\sigma_y^2(1-\rho^2) \\ &D^2(s^2) = \frac{2(n-2)}{n^2}\sigma^4 \end{aligned}\right. \qquad (8)$$

Ainsi l'estimateur s^2 est biaisé. On peut corriger ce biais en prenant un nouvel estimateur s'^2, dont les caractéristiques sont les suivantes :

$$\left\{\begin{aligned} &s'^2 = \frac{n}{n-2}s^2 \\ &E(s'^2) = \sigma^2 = \sigma_y^2(1-\rho^2) \\ &D^2(s'^2) = \frac{2}{n-2}\sigma^4 \end{aligned}\right. \qquad (9)$$

Passons maintenant à l'estimateur m_y. De (7), et de la propriété 1, résulte immédiatement :

$$E(m_y) = E(\bar{y}) - E(b)\,E(\bar{x}-\mu_x) = \mu_y$$

---

(1) Cf. par exemple, Mlle Ullmo, "Dépendance statistique - Liaison stochastique" Cours de l'Inst. de Statistique de l'Université. Paris.

C'est un estimateur sans biais. Utilisant la même propriété 1, nous calculerons sa variance de la manière suivante :

$$D^2(m_y) = D^2(\bar{y}) + E(b^2)\,D^2(\bar{x}) - 2\,E(b)\,E\,[(\bar{x} - \mu_x)(\bar{y} - \mu_y)]$$

$$= \frac{1}{n}\,\sigma_y^2 + \frac{1}{n}\,\sigma_x^2\left[\beta^2 + \frac{\sigma^2}{(n-3)\,\sigma_x^2}\right] - \frac{2\beta}{n}\,\sigma_{xy}$$

$$= \frac{1}{n}\,[\sigma_y^2 - 2\,\beta\,\sigma_{xy} + \beta^2\,\sigma_x^2] + \frac{1}{n\,(n-3)}\,\sigma^2$$

Mais des relations (1) on tire immédiatement :

$$\sigma_x^2 - 2\beta\,\sigma_{xy} + \beta^2\,\sigma_x^2 = \sigma_y^2 - \beta^2\,\sigma_x^2 = \sigma^2$$

D'où finalement les caractéristiques de l'estimateur m_y :

$$\left\{\begin{array}{l} E(m_y) = \mu_y \\ D^2(m_y) = \dfrac{n-2}{n\,(n-3)}\,\sigma^2 = \dfrac{n-2}{n\,(n-3)}\,\sigma_y^2\,(1-\rho^2) \end{array}\right. \qquad (10)$$

IV - ESTIMATION DE LA VARIANCE A PRIORI σ_y^2

Il est utile de pouvoir estimer également la variance à priori σ_y^2. Si, dans la fonction Φ définie en (5), nous remplaçons s^2 par $s_y^2 - b^2\sigma_x^2$, les équations du maximum de vraisemblance, écrites en m_y, s_y et b, conduisent aux mêmes estimateurs b et m_y de β et μ_y que ci-dessus (6), et, pour s_y^2, on obtient simplement :

$$s_y^2 = b^2\,\sigma_x^2 + s^2 \qquad (11)$$

s^2 ayant la même expression qu'en (6). On voit facilement que cet estimateur (11) est biaisé :

$$E(s_y^2) = \sigma_x^2\,[\beta^2 + D^2(b)] + E(s^2)$$

$$= \beta^2\,\sigma_x^2 + \frac{1}{n-3}\,\sigma^2 + \sigma^2 - \frac{2}{n}\,\sigma^2$$

soit, compte tenu de (1)

$$E(s_y^2) = \sigma_y^2 + \frac{6-n}{n\,(n-3)}\,\sigma^2$$

Pour obtenir un estimateur sans biais, on remplacera (11) par une expression de la forme $b^2\,\sigma_x^2 + \lambda\,s^2$. En prenant :

$$\lambda = \frac{n\,(n-4)}{(n-2)\,(n-3)}$$

on obtient l'estimateur sans biais $s_y'^2$:

$$\left\{\begin{array}{l} s_y'^2 = b^2 \ \sigma_x^2 + \dfrac{n\ (n\ -\ 4)}{(n\ -\ 2)\ (n\ -\ 3)}\ s^2 \\ \\ E\ (s_y'^2) = \sigma_y^2 \end{array}\right. \tag{12}$$

Il reste à calculer la variance $D^2\ (s_y'^2)$ de ce dernier estimateur. Utilisant la propriété 1 du paragraphe précédent, nous avons en premier lieu :

$$D^2\ (s_y'^2) = \sigma_x^4\ D^2\ (b^2) + \frac{n^2\ (n\ -\ 4)^2}{(n\ -\ 2)^2\ (n\ -\ 3)^2}\ D^2\ (s^2)$$

Compte tenu de l'expression (8) de $D^2\ (s^2)$, il vient :

$$D^2\ (s_y'^2) = \sigma_x^4\ D^2\ (b^2) + \frac{2\ (n\ -\ 4)^2}{(n\ -\ 2)\ (n\ -\ 3)^2}\sigma^4 \tag{13}$$

Reste à calculer la variance

$$D^2\ (b^2) = E\ (b^4) - E\ (b^2)^2 \tag{14}$$

La valeur de $E\ (b^2)$ résulte de (7). Pour évaluer :

$$E\ (b^4) = E\ [(b\ -\ \beta\ ) + \beta]^4$$

on utilise la propriété 2 du paragraphe précédent. Comme la variable de Student à $n - 1$ degrés de libertés a tous ses moments impairs nuls, et un moment d'ordre 4 égal à $\dfrac{3\ (n\ -\ 1)^2}{(n\ -\ 3)\ (n\ -\ 5)}$, on obtient :

$$\begin{aligned} E\ (b^4) &= E\ [(b\ -\ \beta\ )^4] + 6\ E\ [(b\ -\ \beta)^2] + \beta^4 \\ &= \frac{3}{(n\ -\ 3)\ (n\ -\ 5)}\ \frac{\sigma^4}{\sigma_x^4} + \frac{6\beta^2}{n\ -\ 3}\ \frac{\sigma^2}{\sigma_x^2} + \beta^4 \end{aligned}$$

Reportant ce résultat dans (14), nous obtenons :

$$D^2\ (b^2) = \frac{2\ (n\ -\ 2)}{(n\ -\ 3)^2\ (n\ -\ 5)}\ \frac{\sigma^4}{\sigma_x^4} + \frac{4\beta^2}{n\ -\ 3}\ \frac{\sigma^2}{\sigma_x^2}$$

En substituant dans (13), il vient ainsi :

$$D^2\ (s_y'^2) = \frac{2\ (n\ -\ 4)^2}{(n\ -\ 2)\ (n\ -\ 3)^2}\ \sigma^4 + \frac{4\beta^2}{n\ -\ 3}\ \sigma^2\ \sigma_x^2 + \frac{2\ (n\ -\ 2)}{(n\ -\ 3)^2\ (n\ -\ 5)}\ \sigma^4 \tag{15}$$

L'apparition du facteur $(n - 5)$ au dénominateur du dernier terme de (15) résulte de ce que le moment d'ordre 4 de l'estimateur b n'existe que pour n supérieur à 5.

L'estimateur $s_y'^2$ ne peut donc être utilisé que si $n > 5$. Si n est assez grand, ce dernier terme, qui est en $\dfrac{1}{n^2}$, apparaît comme négligeable. On a, asymptotiquement :

$$D^2\ (s_y'^2) \sim \frac{2}{n}\ \sigma^4 + \frac{4\beta^2}{n}\ \sigma^2\ \sigma_x^2 = \frac{2}{n}\ \sigma^4\ (1\ -\ \rho^2)\ [1\ -\ \rho^2\ +\ 2\ \rho^2]$$

c'est-à-dire :

$$D^2 (s_y'^2) \sim \frac{2}{n} \sigma_y^4 (1 - \rho^4) \tag{16}$$

Il est aisé de faire apparaître cette expression asymptotique, en mettant (15) sous la forme définitive suivante :

$$D^2 (s_y'^2) = \frac{2}{n-3} \sigma_y^4 (1 - \rho^4) - \frac{2(2n^2 - 21n + 46)}{(n-2)(n-5)(n-3)^2} \sigma_y^4 (1 - \rho^2)^2 \tag{17}$$

Le tableau de la page suivante résume les caractéristiques des quatre estimateurs sans biais que nous avons ainsi obtenus dans le cas normal.

V - NOTATIONS DANS LE CAS LOGNORMAL

Supposons, maintenant, que x et y soient des variables lognormales. Les paramètres de leur loi de distribution seront désignés par des lettres grecques, selon les notations suivantes, qui remplacent les notations (1) : (le symbole log représente un logarithme népérien).

$$\left\{\begin{array}{ll} \log \gamma_x & = E(\log x) \\ \log \gamma_y & = E(\log y) \\ \sigma_x^2 & = D^2 (\log x) \\ \sigma_y^2 & = D^2 (\log y) \\ \sigma_{xy} & = E\left[\log \frac{x}{\gamma_x} \log \frac{y}{\gamma_y}\right] \\ \rho & = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \\ \beta & = \rho \frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} \\ \sigma^2 & = \sigma_y^2 (1 - \rho^2) = \sigma_y^2 - \beta^2 \sigma_x^2 \end{array}\right. \tag{18}$$

Les paramètres γ_x et γ_y sont les moyennes géométriques des variables x et y. On sait que les valeurs probables μ_x et μ_y s'en déduisent par les relations lognormales classiques :

$$\left\{\begin{array}{l} \mu_x = E(x) = \gamma_x \exp\left(\frac{\sigma_x^2}{2}\right) \\ \mu_y = E(y) = \gamma_y \exp\left(\frac{\sigma_y^2}{2}\right) \end{array}\right. \tag{19}$$

La droite de régression de log y en log x a pour équation :

$$E[\log y \mid \log x] = \log \gamma_y + \beta \log \frac{x}{\gamma_x}$$

Tableau I

Estimateurs sans biais du cas gaussien

Paramètres	Estimateur sans biais	Variance de cet estimateur
moyenne μ_y	$m_y = \bar{y} - b(\bar{x} \quad \mu_x)$	$D^2(m_y) = \frac{n-2}{n(n-3)} \sigma_y^2 (1 - \rho^2)$
coefficient de régression $\beta = \rho \frac{\sigma_y}{\sigma_x}$	$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$	$D^2(b) = \frac{1}{n-3} \frac{\sigma_y^2 (1 - \rho^2)}{\sigma_x^2}$
Variance liée $\sigma^2 = \sigma_y^2 (1 - \rho^2)$	$s'^2 = \frac{1}{n-2} \sum_i [y_i - \bar{y} - b(x_i - \bar{x})]^2$ $= \frac{1}{n-2} \sum_i (y_i - \bar{y})^2 - \frac{b^2}{n-2} \sum (x_i - \bar{x})^2$	$D^2(s'^2) = \frac{2}{n-2} \sigma_y^4 (1 - \rho^2)^2$
Variance à priori σ_y^2	$s_y'^2 = b^2 \sigma_x^2 + \frac{n-4}{n-3} s'^2$	$D^2(s_y'^2) = \frac{2}{n-3} \sigma_y^4 (1 - \rho^4) - \frac{2(2n^2 - 21n + 46)}{(n-2)(n-5)(n-3)^2} \sigma_y^4 (1 - \rho^2)^2$

Comme la variance liée de log y est σ^2, on en déduit l'équation de la courbe de régression de y en x

$$E\,[y|\,x] = \Upsilon_y \left[\frac{x}{\Upsilon_x}\right] \exp\left[\frac{\sigma_y^2 - \beta^2\,\sigma_x^2}{2}\right]$$

soit, sous forme logarithmique, et compte tenu de (19)

$$\log E\,(y \mid x) = \log \mu_y + \beta \log \frac{x}{\mu_x} + \frac{1}{2}\,\beta\,(1 - \beta)\,\sigma_x^2 \tag{20}$$

En échelle logarithmique, (20) représente encore une droite, mais cette droite ne passe pas par le point central (μ_x, μ_y), sauf dans le cas particulier $\beta = 1$.

Pour estimer les paramètres μ_y et β, et la variance liée σ^2, nous disposons, comme ci-dessus, des valeurs exactes des paramètres Υ_x et σ_x^2 de x, et de n couples expérimentaux $(x_i\ y_i)$. Nous désignerons par m_y, b, s^2 les estimateurs du maximum de vraisemblance de μ_y, β, et σ^2, et de plus nous utiliserons les notations latines suivantes :

$$\left\{\begin{aligned}
\log g_x &= \frac{1}{n}\sum \log x_i \\
\log g_y &= \frac{1}{n}\sum \log y_i \\
S_x^2 &= \frac{1}{n}\sum \left[\log \frac{x_i}{\Upsilon_x}\right]^2 \\
s_x^2 &= \frac{1}{n}\sum \left[\log \frac{x_i}{g_x}\right]^2 = S_x^2 - \left[\log \frac{g_x}{\Upsilon_x}\right]^2 \\
s_{xy} &= \frac{1}{n}\sum \log \frac{x_i}{g_x} \log \frac{y_i}{g_y} = \frac{1}{n}\sum \log \frac{x_i}{\Upsilon_x} \log \frac{y_i}{g_y}
\end{aligned}\right. \tag{21}$$

VI - LES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE (CAS LOGNORMAL)

Le logarithme de la densité de fréquence de (x, y) prend, dans le cas lognormal, la forme suivante :

$$f = -\log 2\pi - \frac{1}{2}\log\sigma_x^2 - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log xy - \frac{1}{2\,\sigma_x^2}\left(\log \frac{x}{\Upsilon_x}\right)^2$$
$$- \frac{1}{2\,\sigma^2}\left[\log \frac{y}{\mu_y} + \frac{1}{2}\,(\sigma^2 + \beta^2\,\sigma_x^2) - \beta \log \frac{x}{\Upsilon_x}\right]^2$$

Nous sommes donc conduits à rechercher le minimum en m_y, b, s^2 de la fonction

$$\Phi\,(m_y,\ s^2,\ b) = n \log s^2 + \frac{1}{s^2}\sum_i \left[\log \frac{y_i}{m_y} + \frac{1}{2}\,(s^2 + b^2\,\sigma_x^2) - b \log \frac{x_i}{\Upsilon_x}\right]^2 \tag{22}$$

En annulant les dérivées en m_y, b et s^2, nous obtenons respectivement :

$$\left\{\begin{array}{l} \log m_y = \log g_y + \frac{1}{2}(s^2 + b^2 \sigma_x^2) - b \log \frac{g_x}{\gamma_x} \\ \sum \left(b\,\sigma_x^2 - \log \frac{x_i}{\gamma_x}\right)\left(\log \frac{y_i}{m_y} + \frac{1}{2} s^2 + \frac{1}{2} b^2 \sigma_x^2 - b \log \frac{x_i}{\gamma_x}\right) = 0 \\ s^2 + \frac{s^2}{n} \sum \left[\log \frac{y_i}{m_y} + \frac{1}{2}(s^2 + b^2 \sigma_x^2) - b \log \frac{x_i}{\gamma_x}\right] = \frac{1}{n} \sum \left[\log \frac{y_i}{m_y} + \right. \\ \qquad \left. \frac{1}{2}(s^2 + b^2 \sigma_x^2) - b \log \frac{x_i}{\gamma_x}\right]^2 \end{array}\right.$$

En portant la valeur de $\log m_y$, donnée par la première équation, dans les deux dernières, on trouve, après des simplifications évidentes :

$$\left\{\begin{array}{l} \log m_y = \log g_y + \frac{1}{2}(s^2 + b^2 \sigma_x^2) - b \log \frac{g_x}{\gamma_x} \\ b = \frac{s_{xy}}{s_x^2} \\ s^2 = \frac{1}{n} \sum \left[\log \frac{y_i}{g_y} - b \log \frac{x_i}{g_x}\right]^2 \end{array}\right. \tag{23}$$

Ainsi les estimateurs b et s^2 se forment, à partir des variables normales $\log x_i$ et $\log y_i$, exactement de la même manière que les estimateurs correspondants (6) du cas normal. Les équations (7) et (8) se transposent donc identiquement, et l'on peut, si l'on veut, former l'estimateur sans biais s'^2 de la variance liée, exactement comme en (9). De même les estimateurs s_y^2 et $s_y'^2$ des équations (11) et (12) se transposent avec toutes leurs propriétés, et en particulier la variance (17).

La seule différence avec le cas normal se manifeste dans le cas de l'estimateur m_y. Il est difficile d'évaluer sa valeur probable exacte. Mais, si n n'est pas trop petit, les estimateurs s^2 et b^2 sont approximativement normaux, et, de leur côté, $\log g_x$ et $\log g_y$ sont des variables gaussiennes. Il en résulte qu'en première approximation log m peut être considéré comme une variable normale - et m_y comme une variable log normale. Comme, en réalité, m_y n'est qu'asymptotiquement lognormal, nous devons supposer n assez grand, et nous contenter des parties principales en $\frac{1}{n}$ des valeurs probables et des variances. Au premier ordre en $\frac{1}{n}$, nous avons :

$$\left\{\begin{array}{l} E\,[\log m_y] = \log \gamma_y + \frac{1}{2} E\,(s_y^2) = \log \mu_y - \frac{1}{2n}\sigma^2 \\ D^2\,[\log m_y] = \frac{1}{n}\,\sigma^2 + \frac{1}{4} D^2\,(s_y^2) = \frac{1}{n}\,\sigma^2 + \frac{1}{2n}\sigma_y^4\,(1 - \rho^4) \end{array}\right.$$

Comme m_y est approximativement lognormal, la valeur probable de m_y s'obtient, à l'approximation en $\frac{1}{n}$, à l'aide de la relation (19), soit :

$$\log E\,(m_y) = E\,\log m_y + \frac{1}{2} D^2\,(\log m_y) = \log \mu_y + \frac{1}{4}\,\sigma_y^4\,(1 - \rho^4)$$

En définitive, l'estimateur m_y est biaisé. Sa valeur probable et sa variance logarithmique sont données, au premier ordre en $\frac{1}{n}$, par :

$$\left\{\begin{aligned} E\,(m_y) &= \mu_y \exp\left[\frac{1}{4n}\sigma_y^4\,(1-\rho^4)\right] \\ D^2 \log m_y &= \frac{1}{n}\sigma_y^2\,(1-\rho^2) + \frac{1}{2n}\sigma_y^4\,(1-\rho^4) \end{aligned}\right. \tag{24}$$

VII - REMARQUE TERMINALE

Les résultats obtenus, c'est-à-dire le tableau I et les formules (24) s'appliquent, en toute rigueur, au cas où les n berlines retenues pour étalonner la droite de correspondance sont choisies au hasard parmi les berlines arrivant au jour. En pratique, ce choix est toujours de nature systématique. Chaque panneau du gisement contribue à la droite de correspondance par un nombre de berlines en rapport avec son tonnage, la règle étant, par exemple, de retenir un nombre fixe de berlines par mètre d'avancement. Et c'est bien ainsi qu'il convient de procéder, puisqu'il est très vraisemblable que l'équilibre radioactif présente un caractère régionalisé, et non aléatoire. Les résultats obtenus ci-dessus ne sont donc pas strictement applicables, mais peuvent être retenus à titre d'approximation : approximation d'ailleurs pessimiste (dans le sens de la sécurité), car on obtiendrait une moins bonne droite en prenant n berlines au hasard, au lieu de les choisir selon un critère systématique.

Ces résultats, peuvent, naturellement, s'appliquer à bien d'autres problèmes pratiques. Citons, par exemple, sans quitter l'industrie minière, l'estimation de la teneur des constituants secondaires d'un minerai, comme l'argent ou l'or liés au plomb ou au cuivre. L'usage est de ne doser pour argent qu'un échantillon de minerai de plomb sur 10 (par exemple), et d'en déduire un coefficient de proportionnalité Ag/Pb. Il s'agit en fait d'une véritable droite de correspondance, et les échantillons analysés pour argent jouent le même rôle que les n berlines retenues pour étalonner la droite.