

Précision sur l'estimation d'une moyenne par une droite de Correspondance.

Le problème abordé ici se pose surtout dans le cas des gisements d'Uranium. Il est, en effet, aisé et peu onéreux de mesurer les radioactivités de toutes les berlines sortant du gisement. Ces radioactivités x sont en corrélation positive - généralement très forte - avec les teneurs y des berlines. 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 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 correspondantes (x_i, y_i) connues expérimentalement. Au contraire, puisque les radioactivités de toutes les berlines sont mesurées, la population des radioactivités x est très bien connue, et les paramètres de leur distribution - soient, dans le cas lognormal, la médiane γ_x et la variance σ_x^2 peuvent être considérés, pour le problème qui nous occupe, comme connus a priori. 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 γ_x et σ_x^2 connus a priori.

Nous supposons que les teneurs y et les radioactivités x sont des variables lognormales de paramètres :

- pour les radioactivités, médiane γ_x variance σ_x^2 (connus a priori)
- pour les teneurs " γ_y " σ_y^2 (à estimer)
- Covariance radioactivité teneur σ_{xy} et coefficient de corrélation ρ
(à estimer)

On introduira aussi le coefficient β de régression de y en x :

$$\sigma_{xy} = \rho \sigma_x \sigma_y = \beta \sigma_x^2$$

Nous désignerons, en outre, par m_y^* le meilleur estimateur de la teneur moyenne m_y , dont nous devons former l'expression à partir des n couples expérimentaux (x_i, y_i) et de γ_x et σ_x^2 . Comme la droite de

correspondance dépend de m_y , β et σ_y^2 , nous aurons aussi à former les estimateurs β^* et σ_y^{2*} . Nous déterminerons ces trois estimateurs par la méthode du maximum de vraisemblance (cf. CRAMER, pp 498 et suiv.)

1°/ Problème normal

Pour des raisons de simplicité, on cherchera d'abord les estimateurs du maximum de vraisemblance $\log \gamma_y^*$, σ_y^{2*} et β^* de la médiane, de la variance et de la pente. La densité de probabilité $f(\log x, \log y)$ du couple de variables normales $f \log x \log y$ a pour logarithme :

$$\text{Log } f = -\log 2\pi - \frac{1}{2} \log \sigma_x^2 \sigma_y^2 (1 - \rho^2) - \frac{1}{2(1-\rho^2)} \left[\frac{\log^2 x / \gamma_x}{\sigma_x^2} - \frac{2\rho \log x / \gamma_x \log y / \gamma_y}{\sigma_x \sigma_y} + \frac{\log^2 y / \gamma_y}{\sigma_y^2} \right]$$

Soit, en remplaçant ρ par $\beta \frac{\sigma_x}{\sigma_y}$

$$(1) \log f = -\log 2\pi - \frac{1}{2} \log \sigma_x^2 - \frac{1}{2} \log (\sigma_y^2 - \beta^2 \sigma_x^2) - \frac{1}{2(\sigma_y^2 - \beta^2 \sigma_x^2)}$$

$$\left[\frac{\sigma_y^2 \log^2 x / \gamma_x}{\sigma_x^2} - 2\beta \frac{\log x / \gamma_x \log y / \gamma_y}{\sigma_x \sigma_y} + \frac{\log^2 y / \gamma_y}{\sigma_y^2} \right]$$

On obtient les estimateurs du maximum de vraisemblance en donnant au paramètre^s inconnus $\log \gamma_y$, β et σ_y^2 les valeurs numériques $\log \gamma_y^*$, β^* et σ_y^{2*} qui rendent maximum la somme

$$\sum_i \log [f(x_i - y_i)]$$

étendue aux n valeurs expérimentales x_i, y_i . Ces trois estimateurs sont les solutions numériques du système :

$$\sum_i \frac{\partial \log f(x_i y_i)}{\partial \log \gamma_y} = \sum_i \frac{\partial \log f(x_i y_i)}{\partial \sigma_y^2} = \sum_i \frac{\partial \log f(x_i y_i)}{\partial \beta} = 0$$

Equation en γ_y .

Elle s'écrit :

$$(2) \quad \sum_i \left[B^* \log \frac{x_i}{\gamma_x} - \log \frac{y_i}{\gamma_y^*} \right] = 0$$

ou

$$\log \gamma_y^* = B^* \log \gamma_x + \frac{1}{n} \sum (\log y_i - B^* \log x_i)$$

Cette équation (2) résoudrait le problème de l'estimation de la médiane γ_y^* si l'on connaissait la valeur vraie de la pente β - et non pas seulement son estimateur B^* .

L'estimateur (2) aurait alors pour variance :

$$D^2 (\log \gamma_y^*) = \frac{1}{n} 6_y^2 (1 - \rho^2)$$

Mais β n'est pas connu a priori. On doit élaborer son estimateur β^* à partir des deux autres équations du maximum de vraisemblance :

Equations en 6_y^2 et β .

Celles-ci nous donnent :

$$(3) \quad \begin{cases} 6_y^{2*} = \beta^{*2} 6_x^2 + \frac{A}{n + \frac{\sum \log x_i / \gamma_x^2}{6_x^2}} \\ B^* \left[\frac{A 6_x^2}{6_y^{2*} - \beta^{*2} 6_x^2} - n 6_x^2 \right] = \sum \log \frac{x_i}{\gamma_x} \log \frac{y_i}{\gamma_y^*} \end{cases}$$

avec

$$(4) \quad A = \sum_i \left[\log \frac{y_i}{\gamma_y^*} - 2 \beta \log \frac{x_i}{\gamma_x} \log \frac{y_i}{\gamma_y^*} + \frac{6_y^{2*}}{6_x^2} \log \frac{x_i}{\gamma_x} \right]$$

En éliminant A des deux équations (3), il vient :

$$(5) \quad B^* = \frac{\sum \log \frac{x_i}{\gamma_x} \log \frac{y_i}{\gamma_y^*}}{\sum \log \frac{x_i}{\gamma_x}}$$

On obtient - en principe- les deux estimateurs simultanés β^* et γ_y^* en résolvant le système (2) - (5), soit en principe pour γ_y^* :

$$(6) \quad \left\{ \begin{aligned} \log \gamma_y^* & \left| 1 - \frac{\left[\sum \log x_i / \gamma_x \right]^2}{n \sum \log x_i / \gamma_x} \right| = \frac{1}{n} \sum \log y_i - \frac{1}{n} \\ & \frac{\sum \log y_i \log \frac{x_i}{\gamma_x}}{\sum \log \frac{x_i}{\gamma_x}} \end{aligned} \right.$$

On voit apparaître dans cette expression les estimateurs :

$$(7) \quad \left\{ \begin{aligned} \log \gamma_x^* &= \frac{1}{n} \sum \log x_i \\ \gamma_x^2 &= \frac{1}{n} \sum \log \frac{x_i}{\gamma_x} \end{aligned} \right.$$

que l'on utiliserait si l'on ne connaissait pas γ_x et γ_x^2 par ailleurs (on note que dans γ_x^2 , c'est la vraie valeur γ_x qui figure et non l'estimateur γ_x^*) (6) peut alors s'écrire :

$$(8) \quad \log \gamma_y^* \left[1 - \frac{\log \gamma_x^*}{\gamma_x^2} \right] = \frac{1}{n} \sum \log y_i - \frac{\log \gamma_x^* / \gamma_x}{\gamma_x^2} \frac{1}{n} \sum \log y_i \log \frac{x_i}{\gamma_x}$$

Tel est l'estimateur qu'il convient d'employer effectivement pour avoir la meilleure estimation de la médiane γ_y . Le calcul de la variance exacte $D^2(\log \gamma_y^*)$ apparaît comme assez coriace. Mais cette variance doit être en $1/n$, et, si n est assez grand, on peut se contenter de sa partie principale en $1/n$, et négliger les termes en $1/n^2$.

Reprenons alors l'estimateur (2)

$$\log \gamma_y^* = \beta^* \log \gamma_x + \frac{1}{n} \sum (\log y_i - \beta^* \log x_i)$$

L'estimateur β^* de β - tel qu'il résulte de la résolution du système (2) - (5) - a lui aussi, une variance en $1/n$. On posera :

$$\beta^* = \beta + \xi$$

β étant la valeur réelle du coefficient, et ξ une variable de valeur moyenne nulle¹ et de variance en $1/n$. Il vient :

1. en réalité de valeur moyenne en $1/n$ - mais ceci ne change rien.

$$\log \gamma_y^* = \frac{1}{n} \sum (\log y_i - \beta \log x_i) + \beta \log \gamma_x - \frac{1}{n} \sum \log \frac{x_i}{\gamma_x}$$

Le dernier terme $\frac{1}{n} \sum \log \frac{x_i}{\gamma_x}$ est le produit de deux variables de moyennes nulles et de variance en $\frac{1}{n}$. Il a lui-même une variance en $\frac{1}{n^2}$, donc négligeable. Il est d'autre part indépendant du premier terme car :

$$\begin{cases} \log \frac{x_i}{\gamma_x} \text{ est indépendant de } \log y_i - \beta \log x_i \\ \sum \log x_i \text{ et } \sum \log y_i \text{ sont indépendants de } \beta^* \end{cases}$$

(d'après les propriétés classiques de la loi normale). Par suite, si on néglige les termes en $1/n^2$, la variance $D^2 (\log \gamma_y^*)$ est encore

$$(9) \quad D^2 (\log \gamma_y^*) = \frac{1}{n} \sigma_y^2 (1 - \rho^2)$$

2°/ Problème lognormal

Ce qui nous intéresse n'est pas la médiane γ_y - que nous savons estimer par γ_y^* mais la moyenne $m_y = \gamma_y e^{\sigma_y^2/2}$ - que nous estimons par

$$(10) \quad m_y^* = \gamma_y^* e^{\sigma_y^{2*}/2}$$

en négligeant le biais² en $1/n$. L'estimateur σ_y^{2*} de la variance, tel qu'il résulte de (3), se met - après quelques manipulations - sous la forme :

$$(11) \quad \sigma_y^{2*} = \beta^{*2} \sigma_x^2 + \frac{1}{n} \sum \left(\log \frac{y_i}{\gamma_y^*} - \beta^* \log \frac{x_i}{\gamma_x} \right)^2$$

- Si β est exactement connu, il suffit d'introduire la variable :

$$\log z = \log y - \beta \log x$$

pour avoir :

$$\sigma_y^{2*} = \beta^2 \sigma_x^2 + \frac{1}{n} \sum \log \frac{z_i}{\gamma_z^*}$$

l'estimateur $\log \gamma_z^*$ étant - d'après (2) - l'estimateur normal

$$\log \gamma_z^* = \frac{1}{n} \sum \log z_i$$

2. Pour éliminer ce biais, il conviendrait de généraliser l'estimateur de SICHEL.

L'estimation de σ_y^2 se ramène à l'estimation de la variance σ_z^2 de z .
Comme on a $\sigma_z^2 = \sigma_y^2 (1 - \rho^2)$, les résultats classiques donnent :

$$(I2) \quad D^2(\sigma_y^{*2}) = \frac{2}{n} \sigma_y^4 = \frac{2}{n} \sigma_y^4 (1 - \rho^2)^2$$

et on note que σ_y^{*2} est indépendant de γ_y^* . D'où - d'après (I0) :

$$(I3) \quad D^2(m_y^*) = \frac{1}{n} \sigma_y^2 (1 - \rho^2) + \frac{1}{2n} \sigma_y^4 (1 - \rho^2)^2$$

- Si β n'est pas connu a priori (c'est le cas général), on se contentera d'une approximation en $\frac{1}{n}$. On pose alors :

$$\begin{cases} \beta^* = \beta + \xi \\ \log \gamma_y^* = \log \gamma_y + \log g \end{cases}$$

ξ et $\log g$ sont des variables de moyenne nulle¹ et de variance en $1/n$.
On écrit (I1) sous la forme :

$$(I4) \quad \sigma_y^{*2} = \beta^2 \sigma_x^2 + \frac{1}{n} \sum \frac{\log^2 z_i}{\gamma_z^2} + 2 \beta \xi \sigma_x^2 - 2 \frac{\xi}{n} \sum \log \frac{x_i}{\gamma_x} \log \frac{z_i}{\gamma_z} \\ - 2 \frac{\log g}{n} \sum \log \frac{z_i}{\gamma_z}$$

ou l'on a négligé les produits de variables telles que ξ et $\log g$. Le produit $\log \frac{z_i}{\gamma_z}$ a une moyenne nulle - puisque x et z sont indépendants. Les

termes $\frac{1}{n} \sum \log \frac{x_i}{\gamma_x} \log \frac{z_i}{\gamma_z}$ et $\frac{1}{n} \sum \log \frac{z_i}{\gamma_z}$ ont donc tous deux moyenne nulle et

variance en $1/n$. Les deux derniers de (I4) peuvent être négligés, comme apportant une contribution en $\frac{1}{n^2}$ à la variance $D^2(\sigma_y^{*2})$ de (I4). On écrira donc :

$$D^2(\sigma_y^{*2}) = \frac{2}{n} \sigma_y^4 (1 - \rho^2)^2 + 4 \beta^2 \sigma_x^4 D^2(\beta^*)$$

La variance $D^2(\beta^*)$ de l'estimateur β^* définie en (5) - à l'approximation en $1/n$ - a son expression classique :

1./ En réalité, il peut y avoir un biais en $\frac{1}{n}$.

$$(I5) \quad D^2(\beta^{\pm}) = \frac{1}{n} \frac{6y^2}{6x^2} (1 - \rho^2)$$

D'où finalement :

$$(I6) \quad D^2(6^{\pm 2}_y) = \frac{2}{n} 6^4_y (1 - \rho^4)$$

et d'après (I0) (car $\gamma^{\pm}_y$ et $6^{\pm 2}_y$ sont encore indépendants)

$$(I7) \quad D^2(m^{\pm}_y) = \frac{1}{n} 6^2_y (1 - \rho^2) + \frac{1}{2n} 6^4_y (1 - \rho^4)$$

3°/ Estimateurs asymptotiquement efficaces.

D'après la méthode indiquée par CRAMER (pp493 et suivantes), nous allons calculer directement les expressions asymptotiques des variances des estimateurs efficaces. L'approximation ainsi faite revient à ne considérer que la partie principale en $\frac{1}{n}$ de ces variances, et à négliger les termes en $\frac{1}{n^2}$ et d'ordre supérieur. Le fait que les expressions ainsi obtenues coïncident avec celles que nous avons obtenues directement dans les deux premiers paragraphes, prouvera que les estimateurs utilisés sont asymptotiquement efficaces - c'est-à-dire qu'il n'est pas possible d'en trouver de meilleurs, à l'approximation du premier ordre.

Méthode.

Si les deux variables x et y obéissent à la loi de distribution à 2 variables $F(xy, abc)$ $dx dy$ dépendant de trois paramètres a, b, c de valeurs numériques inconnues, les estimateurs efficaces $a^{\pm}, b^{\pm}, c^{\pm}$ (s'ils existent) de ces trois paramètres admettent l'ellipsoïde de concentration :

$$n \left[E \left[\frac{\partial \log F}{\partial a} \right]^2 (a^{\pm} - a)^2 + 2 E \left(\frac{\partial \log F}{\partial a} \frac{\partial \log F}{\partial b} \right) (a^{\pm} - a) (b^{\pm} - b) + \dots \right. \\ \left. + E \left(\frac{\partial \log F}{\partial c} \right)^2 (c^{\pm} - c)^2 \right] = 5$$

à partir duquel on obtient les variances $D^2(a^{\pm} - a)$, $D^2(b^{\pm} - b)$ et $D^2(c^{\pm} - c)$

Choix des paramètres à estimer.

On choisira la moyenne m_y des teneurs, la pente $B = \rho \frac{6y}{6x} = \frac{6xy}{6^2_x}$

de la droite de correspondance et la variance 6^2_y .

Pour tenir compte du fait que l'estimateur $m^{\pm}_y$ de m_y sera à peu près lognormal, tandis que les estimateurs $\beta^{\pm}$ et $6^{\pm 2}_y$ seront à peu près normaux, on remplacera m_y par $\log m_y$. Les variances efficaces obtenues

seront $D^2(\log m_y^*)$ - c'est-à-dire la variance logarithmique de m_y^* - et $D^2(\beta^*)$ et $D^2(6_y^{*2})$ - c'est-à-dire les variances arithmétiques des estimateurs β^* et 6_y^{*2} .

Calcul des coefficients de l'ellipsoïde des concentrations.

Exprimée à l'aide des paramètres m_y , β et 6_y^2 , la fonction de probabilité des 2 variables $\log x$ et $\log y$ admet comme log :

$$(I8) \quad \log F = -\frac{1}{2} (6_y^2 - \beta^2 6_x^2) - \frac{1}{2} \left(\log \frac{x}{y} \right)^2 \frac{6_y^2}{6_x^2 (6_y^2 - \beta^2 6_x^2)} \\ + \beta \log \frac{x}{y} \frac{(\log y - \log m_y + \frac{1}{2} 6_y^2)}{6_y^2 - \beta^2 6_x^2} \\ - \frac{1}{2} \frac{(\log y - \log m_y + \frac{1}{2} 6_y^2)^2}{6_y^2 - \beta^2 6_x^2}$$

On posera pour simplifier :

$$(I9) \quad 6_y^2 - \beta^2 6_x^2 = 6_x^2 = 6_y^2 (1 - \beta^2) = D$$

Et on calculera :

$$(20) \quad m_y \frac{\partial \log F}{\partial m_y} = \frac{1}{D} \left(\log \frac{y}{y} - \beta \log \frac{x}{y} \right) \\ \frac{\partial \log F}{\partial 6_y^2} = \frac{1}{2} \frac{1}{D} \left[(-1 + \beta \log \frac{x}{y} - \log \frac{y}{y}) + \frac{1}{D^2} (B^2 \log \frac{x}{y} - 2 \beta \log \frac{x}{y} \log \frac{y}{y} + \log \frac{y}{y}) \right] \\ \frac{\partial \log F}{\partial \beta} = \beta \frac{6_x^2}{D} - \beta \frac{6_y^2}{D^2} \log \frac{x}{y} + \frac{6_y^2 + \beta^2 6_x^2}{D^2} \log \frac{x}{y} \log \frac{y}{y} \\ - \beta \frac{6_x^2}{D^2} \log \frac{y}{y}$$

D'où les moments carrés et rectangulaires :

$$\begin{aligned}
 & \text{IX} \\
 (21) \quad \left\{ \begin{aligned}
 E \left(m_y \frac{\partial \log F}{\partial m_y} \right)^2 &= \iint m_y^2 \frac{(\partial \log F)^2}{\partial m_y} F(xy) dx dy = \frac{1}{D} = a \\
 E \frac{(\partial \log F)^2}{\partial 6^2 y} &= \frac{1}{4D} + \frac{1}{2D^2} = b \\
 E \frac{(\partial \log F)^2}{\partial B} &= \frac{6^2 x (6^2 y + B^2 6^2 x)}{D^2} = c \\
 E \left(m_y \frac{\partial \log F}{\partial m_y} \frac{\partial \log F}{\partial 6^2 y} \right) &= -\frac{1}{2D} = f \\
 E \left(\frac{\partial \log F}{\partial B} m_y \frac{\partial \log F}{\partial m_y} \right) &= 0 = e \\
 E \left(\frac{\partial \log F}{\partial B} \frac{\partial \log F}{\partial 6^2 y} \right) &= -B \frac{6^2 x}{D^2} = d
 \end{aligned} \right.
 \end{aligned}$$

et l'équation de l'ellipsoïde de concentration :

$$n (a x^2 + b y^2 + c z^2 + 2d yz + 2e zx + 2f xy) = s$$

où x , y et z sont mis pour $(\log m^* y - \log m_y)$, $(6^{2*} y - 6^2 y)$, et $(B^* - B)$

Calcul des variances efficaces.

De l'équation de l'ellipsoïde on déduit :

$$(22) \quad \left\{ \begin{aligned}
 D^2 (x) &= \frac{1}{n} \frac{b c - d^2}{abc - cf^2 - be^2 - ad^2 + 2 def} \\
 D^2 (y) &= \frac{1}{n} \frac{ca - e^2}{abc - cf^2 - be^2 - ad^2 + 2 def} \\
 D^2 (z) &= \frac{1}{n} \frac{ab - f^2}{abc - cf^2 - be^2 - ad^2 + 2 def}
 \end{aligned} \right.$$

- X -

des valeurs trouvées (en 2I) pour a b c d e et f on déduit :

$$\left\{ \begin{array}{l} a b c - c f^2 - b e^2 - a d^2 + 2 d e f = 6^2 x 6^2 y (1 - p^2) \\ b c - d^2 = \frac{6^2 x 6^2 y (1 + p^2)}{4 d^3} + \frac{6^2 x 6^2 y (1 - p^2)}{2 D^4} \\ c a - e^2 = \frac{6^2 x 6^2 y (1 + p^2)}{D^3} \\ a b - f^2 = \frac{1}{2 D^3} \end{array} \right.$$

D'où

$$(23) \quad \left\{ \begin{array}{l} D^2 (\log m^* y) = \frac{6^2 y (1 - p^2)}{n} + \frac{6^4 y (1 - p^4)}{2n} \\ D^2 (6^{*2} y) = \frac{2}{n} 6^4 y (1 - p^4) \\ D^2 (B^{\pm}) = \frac{1}{n} \frac{6^2 y}{6^2 x} (1 - p^2) \end{array} \right.$$

Remarque :

Si B est connu a priori (par exemple, si l'on sait a priori que $B = 1$) - en pratique, il ne semble pas que ce soit le cas - l'ellipsoïde est remplacée par l'ellipse de concentration :

$$n (a x^2 + b y^2 + 2 f x y) = 4$$

et l'on a

$$D^2 (x) = \frac{b}{ab - f^2} \quad D^2 (y) = \frac{a}{ab - f^2}$$

Soit

$$\left\{ \begin{array}{l} D^2 (\log m^* y) = \frac{6^2 y}{n} (1 - p^2) + \frac{6^4 y}{2n} (1 - p^2)^2 \\ D^2 (6^{*2} y) = \frac{2}{n} 6^4 y (1 - p^2)^2 \end{array} \right.$$

Les expressions de ces variances efficientes (que B soit ou non connu a priori) sont identiques aux parties principales en $\frac{1}{n}$ des estimateurs déduits du maximum de vraisemblance.

On en déduit que les estimateurs du maximum de vraisemblance, donnés dans les paragraphes 1 et 2 sont bien asymptotiquement efficients. Cela signifie que, pour n pas trop petit, ces estimateurs sont les meilleurs possibles. (D'autre part, le fait de retrouver les mêmes expressions constitue un excellent recoupement, et nous met à l'abri d'une erreur de calcul possible).

4°/ Précision sur une teneur individuelle estimée à partir de sa radio-activité x.

Lorsque la radioactivité x est connue, la teneur y est une variable lognormale de médiane et de variance :

$$(24) \quad \begin{cases} \gamma_{yx} = \gamma_y \left(\frac{x}{\bar{x}} \right)^B \\ \sigma_{yx}^2 = \sigma_y^2 (1 - p^2) \end{cases}$$

Si l'on connaissait les vraies valeurs de γ_y , σ_y^2 et B ces formules résoudraient le problème. La variance de y, en particulier serait $\sigma_y^2(1-p^2)$. Mais on ne connaît pas la vraie médiane $y.x$, mais seulement son estimateur

$$\gamma_{y.x}^* = \gamma_y^* \left(\frac{x}{\bar{x}} \right)^{B^*}$$

Pour une valeur donnée de x, les deux rapports

$$\frac{y}{\gamma_{y.x}} \quad \text{et} \quad \frac{\gamma_{y.x}^*}{\gamma_{y.x}}$$
 sont indépendants

Par suite :

$$D^2 \left(\frac{y}{\gamma_{y.x}^*} \cdot x \right) = D^2 \left(\frac{y}{\gamma_{y.x}} \cdot x \right) + D^2 \left(\frac{\gamma_{y.x}^*}{\gamma_{y.x}} \right) = \sigma_y^2 (1 - p^2) + D^2 \left(\frac{\gamma_{y.x}^*}{\gamma_{y.x}} \right)$$

Calcul de la variance de

$$\frac{\gamma_{y.x}^*}{\gamma_{y.x}} = \frac{\gamma_y^*}{\gamma_y} \left(\frac{x}{\bar{x}} \right)^{B^* - B}$$

On prendra la variance logarithmique de cette expression γ_y^* et B^* étant indépendants (d'après les théorèmes généraux), on a :

$$D^2 \left(\frac{Y_{y.x}^*}{Y_{y.x}} \right) = D^2 (Y_{y.x}^*) + \left(\log \frac{x}{Y_x} \right)^2 D^2 (B^*) = \frac{6^2 y (1-p^2)}{n} + \log \frac{x}{Y_x} \frac{6^2 y (1-p^2)}{n 6^2 x} \quad -2$$

D'où la variance logarithmique donnant la fourchette de y à x fixé autour de $Y_{y.x}^*$:

$$(25) \quad D^2 (Y_{y.x}^*) = 6^2 y (1-p^2) \left[1 + \frac{1}{n} + \frac{1}{n} \frac{\log \frac{x}{Y_x}}{6^2 x} \right] \quad -2$$

Ceci est l'expression asymptotique en $\frac{1}{n}$. D'après Cramer (p.551) l'expression exacte s'obtient en remplaçant n par n - 2.

5°/ Précision sur la valeur probable d'une teneur y à x fixé.

On obtient un résultat notablement différent. La valeur probable de y à x fixé a pour vraie valeur :

$$z = E (y.x) = m_y \left(\frac{x}{Y_x} \right) e^{B - B^2 \frac{6^2 x}{2}}$$

On l'estime par l'estimateur approximativement lognormal z^* défini par :

$$(26) \quad \log z^* = \log m_y^* + B^* \log \frac{x}{Y_x} - B^{*2} \frac{6^2 x}{2}$$

On note que m_y^* et B^* sont indépendants $\left[\text{car } e = E \left(\frac{\partial \log F}{\partial B} m_y \frac{\partial \log F}{\partial m_y} = 0 \right) \right]$

D'où la variance logarithmique de z^* :

$$D^2 (z^*) = D^2 (m_y^*) + D^2 \left(B^* \log \frac{x}{Y_x} - B^{*2} \frac{6^2 x}{2} \right)$$

On posera $B^* = B + \xi$ ξ ayant moyenne nulle et variance en $\frac{1}{n}$

d'où $B^{*2} = B^2 + 2 \xi B + \xi^2$ on négligera ξ^2 comme ayant une variance en $\frac{1}{n^2}$

et on prendra $B^{*2} = B^2 + 2B (B^* - B) = 2 B B^* - B^2$

d'où

$$D^2 \left[B^* \log \frac{x}{Y_x} - B^{*2} \frac{6^2 x}{2} \right] = D^2 \left[B^* \left(\log \frac{x}{Y_x} - B \frac{6^2 x}{2} \right) + B^2 \frac{6^2 x}{2} \right] =$$

$$= (\log \frac{x}{\bar{y}} - B 6^2 x)^2 D^2(B^*)$$

D'où finalement :

$$(26) \quad D^2(z^*) = \frac{6^2 y (1 - p^2)}{n} + \frac{6^4 y (1 - p^4)}{2 n} + \frac{1}{n} (\log \frac{x}{\bar{y}} - B 6^2 x)^2$$

$$\frac{6^2 y}{6^2 x} (1 - p^2)$$

Il s'agit ici encore d'une expression asymptotique en $\frac{1}{n}$. On en déduit la fameuse hyperbole de confiance $z^* \pm 2 D(z^*)$ qui donne l'erreur sur l'estimation de la valeur probable d'une teneur lorsque la radioactivité est connue.

On notera la signification très différente de (25) et (26). De (25), on déduit la fourchette de l'erreur commise en déduisant la teneur y d'une berline de sa radioactivité et de la droite de correspondance. La partie principale de la variance d'erreur, $6^2 y (1 - p^2)$ est finie. L'équation (26) donne l'erreur commise en assimilant, au point d'abscisse x , la droite de correspondance réelle (inconnue) à la droite déterminée à partir des n couples (x_i, y_i) . La variance est en $1/n$ et tend vers 0 lorsque n augmente indéfiniment.

6°/ Remarques terminales.

I/ Nous avons admis, implicitement, que l'échantillonnage était aléatoire. Les formules données ci-dessus s'appliquent, en toute rigueur, au cas où les n berlines retenues pour étalonner la droite de correspondance sont choisies au hasard parmi toutes les berlines arrivant au jour. En pratique, l'échantillonnage est toujours de nature systématique. Chaque panneau contribue à la droite de correspondance par un nombre de berlines en rapport avec son tonnage, la règle étant par exemple de retenir une berline pour tant de mètres d'avancement. Et c'est bien ainsi qu'il convient de procéder, puisqu'il est très vraisemblable que l'équilibre radioactif possède un caractère régionalisé - et non pas aléatoire. Les formules données ci-dessus ne sont donc pas strictement applicables, mais peuvent être retenues à titre d'approximation. Cette approximation est d'ailleurs pessimiste (dans le sens de la sécurité), car il est bien certain qu'on aurait une moins bonne droite de correspondance en prenant

- XIV -

n berlines au hasard, qu ^{le u} bien d'en choisir le même nombre selon des critères systématiques.

2/ Les formules ci-dessus peuvent s'appliquer aussi à l'estimation de la teneur des constituants secondaires d'un minéral, par exemple l'argent ou l'or liés au plomb ou au cuivre. L'usage est, de ne doser pour Ag qu'un échantillon de minéral 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 Ag/Pb, et les échantillons analysés pour Ag jouent le même rôle que les n berlines retenues pour l'établissement de la droite de correspondance.

--:--:--:--:--:--:--:--:--