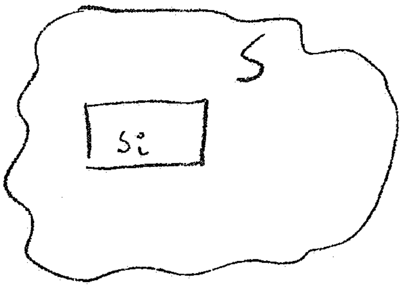
Février 1959

NOTE STATISTIQUE N° 18

Les échantillonages systématiques hétérogènes

I – Assimilation d'un échantillon systématique à un échantillon aléatoire.



Soit un échantillon s de teneur x destiné à échantillonner le panneau S_i de teneur m_i dans un gisement S. Si S_i est le gisement lui-même, on peut toujours l'imaginer comme une portion d'un gisement fictif S beaucoup plus grand.

L'échantillon s pourra être constitué de l'ensemble de sondages à maille rigide, ou de galeries et recoupes etc... Dans tous les cas, il présente un caractère systématique.

Nous admettrons que s peut être considéré comme implanté au hasard dans S_i, si la covariance $\sigma_{x m_i}$ de l'échantillon s et du panneau S_i dans S peut être considérée comme égale à la variance $\sigma^2_{m_i}$ de S_i dans S, c'est à dire si les deux intégrales

$$\frac{1}{S_i s}\iint_s ds \iint_{S_i} C_{x_{12}}\, dS \qquad \text{et} \qquad \frac{1}{S_i^2}\iint_{S_i} ds \iint_{S_i} C_{x_{12}}\, dS$$

(1 dans s et deux dans S_i) (1 et 2 dans S_i)

sont équivalentes.

Tel sera le cas pour une maille rigide de sondages, un réseau régulier de traçages ou de montages, etc... mais non pas pour un sondage unique implanté au centre du panneau, une galerie centrale unique, un échantillonnage périphérique ...

Si j'ai reconnu deux parties d'un gisement par deux échantillonnages homogènes de type différent, l'un par galeries et recoupes et l'autre par sondages à maille rigide par exemple, les deux variables $\frac{x_1}{m_1}$ et $\frac{x_2}{m_2}$ sont indépendantes, ou si l'on veut, les deux covariances $\sigma_{x_1 x_2}$ et $\sigma_{m_1 m_2}$ sont égales. Mais ni x_1 ni x_2 ne pourront être considérés comme implantés au hasard dans S.

II – Estimation de la teneur moyenne par l'artifice du gisement infiniment grand.

Un échantillon s dans le gisement S_i de teneur moyenne réelle m_i donne une estimation x avec une variance d'échantillonnage (d'extension) σ_i^2. s est par exemple un réseau de traçages et montages

.../...

dont x est la teneur moyenne.

On peut imaginer que m_i est un panneau d'un gisement fictif très grand S de teneur m. Reprenant le raisonnement de Krige, on voit qu'à m_i fixé x a pour valeur probable m_i et pour variance σ_i^2. Mais inversement, et c'est le problème pratique intéressant, une fois x connu, m_i apparaît comme une variable lognormale de moyenne:

$$E(m_i \cdot x) = e^{\frac{\sigma_i^2 \sigma_\gamma^2}{2(\sigma_i^2+\sigma_\gamma^2)}} \, x^{\frac{\sigma_\gamma^2}{\sigma_i^2+\sigma_\gamma^2}} \, m^{\frac{\sigma_i^2}{\sigma_i^2+\sigma_\gamma^2}}$$

$$= x^{\frac{\sigma_\gamma^2}{\sigma_i^2+\sigma_\gamma^2}} \left[m\, e^{\sigma_\gamma^2/2} \right]^{\frac{\sigma_i^2}{\sigma_i^2+\sigma_\gamma^2}}$$

et de variance

$$D^2(m_i \cdot x) = \frac{\sigma_i^2 \sigma_\gamma^2}{\sigma_i^2+\sigma_\gamma^2}$$

Si l'on suppose S infiniment grand, son influence s'évanouit car $\sigma_\gamma^2 \to \infty$ et il reste

$$\begin{cases} E(m_i \cdot x) = e^{\sigma_i^2/2}\, x \sim \text{indéterminé, mais prendre } x \\ D^2(m_i \cdot x) = \sigma_i^2 \end{cases}$$

pas de sens

~~Tout se passe comme si, à x connu, m_i était une variable lognormale de médiane x et de variance σ_i^2. Il est un peu surprenant de voir que la valeur probable de m_i est toujours supérieure à x. Cependant, le facteur $e^{\sigma_i^2/2}$ est toujours petit et peut être négligé. De plus, l'intervalle de sécurité au seuil de 95 % est~~

~~$x\, e^{-2\sigma_i}$ $\qquad x\, e^{2\sigma_i}$~~

~~C'est en pratique la valeur inférieure $x\, e^{-2\sigma_i}$ qui doit être utilisée.~~

III - Les échantillonnages systématiques hétérogènes .-

Il s'agit d'échantillonnages tels que la densité des prélèvements soit différente dans les différentes parties du gisement sans que le schéma d'ensemble puisse être considéré comme aléatoire. Par exemple : une partie du gisement a été échantillonnée par sondages systématiques à une certaine maille, l'autre par travaux miniers ~~mission~~ ou par sondages à une maille différente etc .....

Equations générales (méthode du gisement infiniment grand)

Pour poser le problème dans toute sa généralité, supposons qu'un gisement G, de teneur moyenne réelle z ait été échantillonné par n échantillons

...//...

$S_1 \ldots S_n$ de caractéristiques différentes, la forme et la position de chacun d'eux étant bien déterminées. Soit x_i la teneur de l'échantillon S_i. Par un artifice de l'esprit, on peut toujours imaginer que le gisement S soit un panneau d'un gisement S' de teneur m. On verra qu'en supposant S' infiniment grand, la valeur de m perd toute influence.

Soient alors dans le gisement fictif S'

$\sigma^2_{x_i}$ la variance de l'échantillon S_i

σ^2_{z} la variance du panneau S équivalent au gisement réel.

$\sigma_{x_i x_j}$ et $\sigma_{x_i z}$ les covariances de l'échantillon S_i avec l'échantillon S_j et le panneau S lorsque les dispositions relatives de S_i, S_j et S sont celles qui nous sont imposées.

Ces variances et covariances sont toutes calculables par intégration de L r, à une même constante A près qui ne dépend que de S' et tend vers l'infini en même temps que S'.

En toute rigueur, nous devrions déterminer les lois de répartition de la variable lognormale z liées par la fixation successive de x_1, x_n, et voir ce qu'il advient des formules obtenues pour S' infinie. Ce mode de calcul serait particulièrement pénible. En fait, nous savons qu'il conduit, pour la valeur probable de z lorsque les n valeurs x_i sont fixées, à une expression de la forme

$$(1) \qquad u = E(z, x_1, x_2 \ldots x_n) = C\, m^{a_0} \prod_{i=1}^{n} x_i^{a_i} ;$$

où C est une constante, et où l'exposant a_0, comme on peut le voir immédiatement par recurrence, serait égal à

$$(2) \qquad a_0 = 1 - \sum_{i=1}^{n} a_i$$

Nous considérerons donc l'expression de u définie par l'équation (1) comme un estimateur de la variable aléatoire z. u devient alors elle-même une variable lognormale, dont les propriétés dépendent des valeurs attribuées aux coefficients a_i. Et nous déterminerons ces coefficents a_i par les deux conditions suivantes :

1°/ La valeur probable de z à u fixé, $E(z.u)$ doit être égale à u elle même (à une constante près aisément calculable)

2°/ la variance de z à u fixé doit prendre la valeur la plus petite possible.

Si nous désignons par σ_u^2 et σ_{uz} la variance et la covariance avec z dans S' de la variable aléatoire u définie par l'équation (I), la valeur probable de z à u fixé sera de la forme

(3) $$E(z.u) = c' \, u^{\frac{\sigma_{uz}}{\sigma_u^2}} \, m^{1-\frac{\sigma_{uz}}{\sigma_u^2}} \qquad (c' = c\,\tfrac{tr}{\,})$$

et la variance de z à u fixé

(4) $$D^2(z.u) = \sigma_z^2 - \frac{\overline{\sigma_{uz}}^2}{\sigma_u^2}$$

La première condition s'exprimera par :

(5) $$\frac{\sigma_{uz}}{\sigma_u^2} = 1$$

et la deuxième en écrivant que l'expression (4) est minimum. Explicitons les valeurs de σ_u^2 et σ_{uz} en fonction des coefficients a_i. L'équation (1) permet d'écrire :

(6) $$\begin{cases} \sigma_u^2 = \sum_{ij} a_i a_j \sigma_{x_i x_j} \\ \sigma_{uz} = \sum_i a_i \sigma_{z x_i} \end{cases}$$

lorsque S' augmente indéfiniment, les $\sigma_{x_i x_j}$ et les $\sigma_{z x_i}$ sont majorés d'une même constante A qui tend vers l'infini. A la limite, donc, la condition (5) se réduit à :

$$\frac{\left(\sum_i a_i\right)^2}{\sum a_i} = 1$$

avec comme seule solution admissible :

(7) $$\sum_i a_i = 1$$

Remarquons tout de suite que le terme en m disparait alors de l'expression de u, comme il résulte de l'équation (2). On verra de même que, pour S' infini, l'expression (4) de la variance de z à u fixé se réduit à :

$$(8)\qquad D^2(z.u) = \sigma_u^2 + \sigma_z^2 - 2\sigma_{uz}$$

expression indépendante de S', et du reste égale à la variance de la variable z/u

L'expression (6) s'explicite :

$$(9)\qquad D^2(z.u) = \sum_{ij} a_i a_j \sigma_{x_i x_j} + \sigma_z^2 - 2\sum_i a_i \sigma_{z x_i}$$

Les a_i sont déterminés par la condition de rendre (9) minimum compte tenu de (7) ce qui s'exprime par le système d'équations :

$$(10)\qquad \begin{cases} \sum_i a_i = 1 \\ \sum_j a_j \sigma_{x_i x_j} - \sigma_{z x_i} = \lambda \qquad (\text{pour } i = 1, 2 \ldots n) \end{cases}$$

C'est un système linéaire de n + 1 équations à n + 1 inconnues : les n coefficients a_i et le multiplicateur de Lagrange λ.

Quant aux termes $\sigma_{z x_i}$ on pourra, dans la plupart des cas pratiques, considérer qu'ils ont même valeur pour tous les x_i. En effet, si l'échantillon S_i au lieu d'être implanté dans une position bien déterminée dans S, y était implanté au hasard tous les $\sigma_{z x_i}$ seraient égaux à σ_z^2 variance de S dans S'. Ils n'en diffèrent, en fait, que par des termes de l'ordre de la dispersion absolue σ, termes représentant l'élément préférentiel de la corrélation de S_i et de S imputable à la position privilégiée de S_i dans S. Cependant le multiplicateur de Lagrange λ est du même ordre de grandeur, comme on peut le voir à partir des équations (9) et (10), et par suite ces termes ~~ne peuvent,~~ théoriquement, être négligés. Mais, en pratique, les échantillons S_i seront, chacun pour leur part, le résultat d'échantillonnages homogènes d'un petit nombre de panneaux $\neq$ de S dont aucun n'occupera de position privilégiée, sauf géométrie aberrante de S. S'agissant de sous-échantillonnages homogènes, les covariances $\sigma_{x_i z}$ seront égales aux covariances des panneaux et de S, elles-mêmes égales entre elles puisque les panneaux sont en position équivalente (Il serait bon, cependant, de le vérifier sur quelques exemples numériques). Les équations (10) se réduisent alors à :

$$\sum_i a_i = 1$$

(11) $$\sum_j a_j \sigma_{x_i x_j} = \lambda' \qquad (i = 1, 2 \ldots n)$$

$\sigma_{x_i x_j}$ représentant cette fois la covariance des échantillons S_i et S_j dans le gisement réel S.

<u>Equations générales (méthodes du maximum de vraisemblance)</u>

Nous raisonnerons maintenant dans le gisement réel S dont nous cherchons à estimer la teneur réelle m, sans passer par l'intermédiaire d'un grand gisement fictif S'. Les échantillons S_i ont, dans S, les variances $\sigma^2_{x_i}$ et les covariances $\sigma_{x_i x_j}$. Relevons au passage qu'admettre que S_i a une variance définie $\sigma^2_{x_i}$ dans S revient à supposer que S_i peut être supposé implanté au hasard dans S, ou du moins n'y occupe pas de position privilégiée. Cependant, nous ne supposons pas que les $\sigma_{x_i x_j}$ soient tous nuls dans S. Celà revient à peu près à admettre que la figure indéformable constituée par l'ensemble des S_i dans leurs positions relatives imposées est implantée en bloc dans S de façon aléatoire. Cette hypothèse est, en fait, équivalente à celle qui a été faite au paragraphe précédent lorsqu'il a été admis que les covariances $\sigma_{x_i j}$ avaient toutes la même valeur. La notion de variance dans S ne peut avoir de sens que pour des échantillons n'occupant pas de position privilégiée dans S. Il ne faudra donc pas s'étonner que le raisonnement de ce paragraphe conduise aux équations (11).

La fonction de probabilité a priori des variables x_i est de la forme :

(12) $$f(x_1, x_2 \ldots x_n ; m) = e^{-\sum_{ij} L\frac{x_i}{\gamma_i} L\frac{x_j}{\gamma_j} C_{ij}}$$

où les C_{ij} représentent les composantes de la matrice inverse de la matrice de terme général $\sigma_{x_i x_j}$, et où les γ_i sont les médianes des x_i. l'exposant se met sous la forme

(13) $$\sum_{ij} (Lx_i - m)(Lx_j - m) C_{ij} \quad + \quad \text{terme indépendant de m}$$

L'estimateur de m qui donne le maximum de vraisemblance est celui qui maxime l'expression (12) ou minimise l'expression (13) Sa valeur est donc définie par l'équation :

$$\sum_{ij} C_{ij} \left[Lx_i + Lx_j - 2Lm \right] = 0$$

Soit

.....

$$Lu = \sum_i a_i x_i$$

$$a_i = \frac{\sum_j C_{ij}}{\sum_{ij} C_{ij}}$$

On est ainsi conduit à un estimateur de la moyenne de la forme

(15) $$u = \prod_i x_i^{a_i}$$

avec, comme au paragraphe précédent, la condition

(7) $$\sum a_i = 1$$

L'équation (14) est d'ailleurs équivalente au système (11). En effet les Cij représentent la matrice inverse de la matrice des $\sigma_{x_i x_j}$. Le terme Cij est donc proportionnel au coefficient de $\sigma_{x_i x_j}$ dans le développement du déterminant des $\sigma_{x_i x_j}$. Par suite, l'expression $\sum_j C_{ij}$ est proportionnelle à la valeur que prend ce déterminant lorsque les éléments de la ième ligne sont remplacés par l'unité. L'équation (14) représente la solution de Cramer du système (11)

IV - Exemples numériques.-

Soit un gisement S comprenant deux panneaux S_1 et S_2 échantillonnés de façon différente. Ces échantillonnages ont conduit aux estimations x et y des teneurs de S_1 et S_2. Soient σ_x^2 et σ_y^2 les variances dans S des estimateurs x et y. σ_x^2 par exemple, est la somme de la variance d'extension de x dans S_1 et de la variance de S_1 dans S. Soit enfin σ_{xy} la covariance de x et y que l'on prendra en principe égale à la covariance des panneaux S_1 et S_2. On cherche un estimateur de la forme

(16) $$u = A x^a y^{1-a}$$

et de variance minima. La variance de u est :

(17) $$\sigma_u^2 = a^2 \sigma_x^2 + (1-a)^2 \sigma_y^2 + 2a(1-a)\sigma_{xy}$$

Elle est minima pour

(18) $$a = \frac{\sigma_y^2 - \sigma_{xy}}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}}$$

Avec cette valeur de a_1 la variance est égale à :

(19) $$\sigma_u^2 = \frac{\sigma_x^2\,\sigma_y^2 - \sigma_{xy}^2}{\sigma_x^2+\sigma_y^2-2\sigma_{xy}}$$

On détermine enfin la constante A en écrivant que la valeur probable de l'estimateur u est égale à m . On obtient facilement la condition :

(20) $$A = e^{\frac{(\sigma_x^2-\sigma_{xy})(\sigma_y^2-\sigma_{xy})}{2(\sigma_x^2+\sigma_y^2-2\sigma_{xy})}}$$

On remarquera que les coefficients a de la formule (18), comme les a_i des formules (11) et (14) sont indépendants de la dispersion absolue α. Ce sont des facteurs purement géométriques, calculables par intégration de Ir. Seule la constante A dépend de la dispersion absolue.

Ier exemple :

Soit un gisement stratiforme de surface $S = 10^6$, divisé en deux panneaux égaux S_1 et S_2. S_1 a été reconnu par 100 sondages et S_2 par 25 sondages à maille rigide. La section des carottes est 10^{-2}. On admet que la formule ordinaire de de Wijs est valable, la dispersion absolue étant α. Soit x la moyenne des 100 sondages de S_1 et y la moyenne des 25 sondages de S_2. On a

$$\sigma_x^2 = \alpha\left[\frac{1}{100}\left(L\frac{10^6}{2\times 10^2\times 10^{-2}} - \frac{1}{2}\right) + L2\right] = \alpha\,\frac{81.41}{100}$$

$$\sigma_y^2 = \alpha\left[\frac{1}{25}\left(L\frac{10^6}{2\times 25\times 10^{-2}} - \frac{1}{2}\right) + L2\right] = \alpha\,\frac{124.96}{100}$$

$$\sigma_{xy} = -\alpha L2 = -\alpha\,\frac{69}{100}$$

Les formules 18 donnent

$$\begin{cases} a = 0.565 \\ 1-a = 0.435 \end{cases} \qquad \begin{cases} \sigma_u^2 = 0.158\,\alpha \\ A = e^{0.425\alpha} \end{cases}$$

On prendra comme estimateur

$$u = e^{0.475\alpha}\, x^{0.565}\, y^{0.435}$$

Il est peut être surprenant d'affecter des poids différents à deux parties égales d'un même gisement. La raison en est que x n'est pas seulement un estimateur de S_1, mais aussi, à un moindre degré, de S_2 et qu'en fait x est un meilleur estimateur de S_2 que ne l'est y de S_1. D'où une influence plus grande attribuée à x.

Si l'on attribuait la même influence à x et y, la variance de l'estimateur correspondant

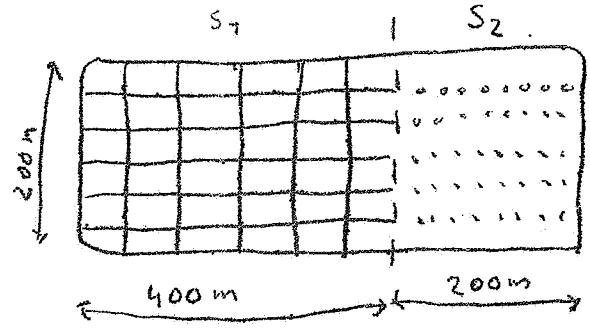
$$\sigma = A' x^{0.5} y^{0.5}$$

serait

$$\frac{1}{4}\left[\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}\right] = 0.17\,\alpha$$

soit 0,17 α au lieu de 0,158 α. La perte de précision n'est pas considérable, et on pourrait accepter, en pratique, d'accorder la même influence à x et y. On n'a d'ailleurs le droit d'utiliser l'estimateur u que si le panneau S_1 a été choisi au hasard. Si, par exemple, la doctrine de recherche est de reconnaître d'abord à large maille l'ensemble du gisement, puis de resserrer la maille sur la partie qui paraît la plus riche, l'estimateur u introduirait une surestimation systématique des réserves : il faut alors utiliser un estimateur du type σ (ou du type $\frac{x+y}{2}$ du reste pratiquement équivalent pour ces faibles dispersions.)

2ème exemple :



Soit un gisement filonien S formant rectangle 600 m x 200 m reconnu par 5 niveaux distants de 40 m. Dans la partie gauche S_1 (rectangle 200 x 400) les 5 niveaux ont été tracés au minerai et de plus 5 montages, distants de 80 m, ont été également échantillonnés. Dans la partie droite, au contraire qui forme un carré de 200 m. de côté, les niveaux ont été creusés en roche, et le filon a été reconnu par sondages horizontaux tous les 10 mètres, soit 20 sondages par niveaux : le panneau S_2 est en somme reconnu par 100 sondages implantés à maille rectangulaire 10 x 40

Le problème posé est d'estimer les teneurs de S_1, S_2 et S et de déterminer la précision de ces estimations.

A/ Estimation du panneau S_1 -

Soit x_1 la teneur moyenne des 5 traçages (2000 mètres) et x_2 la teneur moyenne des 5 montages (1000 mètres). Les deux variables x_1 et x_2, de variances $\sigma^2_{x_1}$ et $\sigma^2_{x_2}$ dans S_1 constituent deux estimateurs indépendants de la teneur moyenne x du panneau S_1, car leur covariance dans S_1 est manifestement négligeable. On doit donc, en principe, prendre comme estimateur de x une expression de la forme

$$
u = A x_1^{a} x_2^{1-a}
$$

$$
(21) \qquad a = \frac{\sigma^2_{x_2}}{\sigma^2_{x_1} + \sigma^2_{x_2}}
$$

$$
\sigma^2_u = \frac{\sigma^2_{x_1}\,\sigma^2_{x_2}}{\sigma^2_{x_1} + \sigma^2_{x_2}}
$$

1ère approximation :-

Supposons que les variances soient données par les formules d'approximations

$$
(22) \qquad \sigma^2_{x_1} = \alpha \frac{\pi}{2} \frac{S_1}{L_1^2} \qquad\qquad \sigma^2_{x_2} = \alpha \frac{\pi}{2} \frac{S_1}{L_2^2}
$$

L_1 = longueur totale des traçages

L_2 = longueur totale des montages/

On a alors :

$$
(23) \qquad \begin{cases} a = \dfrac{L_1^2}{L_1^2 + L_2^2} \\[2ex] \sigma^2_u = \alpha \dfrac{\pi}{2} \dfrac{S_1}{L_1^2 + L_2^2} \end{cases}
$$

La bonne règle, dans l'hypothèse d'approximation adoptée ici, est de pondérer traçages et montages par les carrés de leur longueur. La variance de l'estimateur est alors donnée par une formule du type (22) où les longueurs des traçages et des montages ont été composées suivant la règle de Pythagore.

Numériquement :

$$(24)\quad \begin{cases} \sigma^2_{x_1} = \alpha \frac{\pi}{2} \frac{S_1}{L_1^2} = \alpha \frac{\pi}{2} \frac{8.10^4}{4.10^6} = \frac{3.14}{100} \alpha \\ \sigma^2_{x_2} = \alpha \frac{\pi}{2} \frac{S_2}{L_2^2} = \alpha \frac{\pi}{2} \frac{8.10^4}{10^6} = \frac{12.56}{100} \alpha \end{cases}$$

$$(25)\quad \begin{cases} a = 0.8 \\ 1-a = 0.2 \end{cases}$$

$$(26)\quad \begin{cases} \sigma^2_u = \frac{2.51}{100} \alpha \\ A = e^{\sigma^2_u/2} \end{cases}$$

Le terme de biais A est négligeable. On prendra

$$(27)\quad u = x_1^{0.8} \, x_2^{0.2}$$

<u>2ème approximation.-</u>

On peut craindre, au moins pour les montages, que $\operatorname{tg} \phi = \frac{80}{200} = 0.4$ n'ait une valeur un peu forte pour l'approximation du 1er degré que constituent les formules (22). Utilisant les formules du rectangles, on obtient les valeurs exactes suivantes

$$(28)\quad \begin{cases} \sigma^2_{x_1} = \frac{3.370}{100} \alpha \\ \sigma^2_{x_2} = \frac{10.824}{100} \alpha \end{cases}$$

que l'on comparera utilement aux valeurs (24). On a ainsi :

$$(29)\quad \begin{cases} a = 0.765 \\ 1-a = 0.235 \\ \sigma^2_u = \frac{2.58}{100} \alpha \end{cases}$$

.../...

Le meilleur estimateur est donc :

$$(30) \qquad u = x_1^{0.765} \, x_2^{0.235}$$

Remarque : On aurait pu, x_1 et x_2 étant deux estimations indépendantes de x_1 juger intuitivement qu'il fallait prendre leur moyenne arithmétique $\frac{x_1 + x_2}{2}$, ou, ce qui revient à peu près au même, leur moyenne géométrique $\sqrt{x_1 x_2}$. La variance de l'estimateur serait dans ce cas :

$$(31) \qquad \frac{\sigma^2_{x_1} + \sigma^2_{x_2}}{2} = \frac{3.73}{100}\,\alpha$$

L'estimateur $\frac{x_1 + x_2}{2}$ est donc, en fait, moins bon que x_1 tout seul.

Plus naturellement encore, on aurait pu prendre la teneur moyenne réelle de l'ensemble des traçages et des montages, soit

$$\frac{L_1 x_1 + L_2 x_2}{L_1 + L_2} = \frac{2\,x_1 + x_2}{3}$$

La variance de l'estimateur est alors

$$\frac{L_1^2 \sigma^2_{x_1} + L_2^2 \sigma^2_{x_2}}{(L_1 + L_2)^2} = \frac{2.70}{100}\,\alpha$$

elle est assez voisine de la variance de l'estimateur (30) mais encore supérieure. Du point de vue pratique, cependant, la différence est assez faible pour que l'on puisse utiliser sans trop de remords un tel estimateur.

En pratique, du reste, on n'aura généralement pas le loisir de calculer les variances exactes par la formule du rectangle, et on devra se contenter des formules d'approximation du type (24), qui conduisent à pondérer traçages et montages par les carrés de leurs longueurs. La formule (27) avec les valeurs (28) des variances, donne une variance d'échantillonnage :

$$\overline{0.8}^2 \sigma^2_{x_1} + \overline{0.2}^2 \sigma^2_{x_2} = \frac{2.63}{100}\,\alpha$$

également un peu plus forte que l'optimum. Les deux estimateurs obtenus en pondérant x_1 et x_2 par les longueurs ou par les carrés des longueurs sont assez comparables, et il serait sans doute utile, en pratique, de les calculer tous les deux.

Pour ce qui suit, nous adopterons l'estimateur (30) de la teneur x de S_1.

.../...

B - <u>Estimation du panneau S_2</u>

Les 100 sondages étant tous équivalents, l'estimateur de la teneur y de S_2 est l'estimateur lognormal habituel. La difficulté est ici de calculer la variance de cet estimateur, qui nécessite théoriquement le calcul des 4950 covariances des sondages deux à deux. Il est préférable de recourir à l'approximation ci-après, qui n'introduira que des erreurs négligeables. Si les teneurs réelles des 5 traçages étaient connues, la variance serait celle des formules du rectangle, soit à peu près $\alpha \frac{\pi}{2} \frac{s_2}{L^2}$

Cependant, nous ne connaissons pas la teneur réelle du niveau 1, par exemple, mais seulement une estimation de celle-ci à partir de 20 sondages implantés tous les 10 mètres. D'où une erreur sur l'estimation de chaque niveau. Ces différentes erreurs, sur chaque niveau, sont du reste indépendantes les unes des autres car, par exemple, la covariance moyenne des 20 sondages du N 1 et des 20 sondages du N2 est presque rigoureusement égale à la covariance de ces deux niveaux. Par suite, si σ_i^2 désigne la variance de l'estimation d'un niveau par ses 20 sondages, la variance d'échantillonnage totale est :

$$(32) \qquad \sigma_E^2 = \frac{\sigma_i^2}{5} + \alpha \frac{\pi}{2} \frac{s_2}{L^2}$$

<u>Variance σ_i^2</u> Pour estimer la teneur d'un niveau, nous prenons l'estimateur :

$$A \left(y_1 \, y_2 \ldots \, y_{20} \right)^{1/20}$$

où les y_i représentent les teneurs des 20 sondages. La variance σ_i^2 correspondante est :

$$(33) \qquad \sigma_i^2 = \frac{20\, \sigma_{y_i}^2 + 2 \sum \sigma_{y_i y_j}}{400}$$

$\sigma_{y_i}^2$ et $\sigma_{y_i y_j}$ étant les variances et covariances des sondages dans le traçage

On sait que la variance $\sigma_{y_i}^2$ est la même que celle d'un échantillon linéaire de longueur d double du diamètre de la carotte. Prenant comme unité la distance 10 m de deux sondages consécutifs, on calculera la valeur moyenne de σ^2 sur l'ensemble des sondages du niveau :

$$\underset{\text{sondages}}{E(L\eta)} = \frac{20(L d - 1.5)}{400} + \frac{2}{400}\left[L 19 \ldots + L 18 \ldots + \ldots + 2 L 8 + L 19 \right]$$

$$= \frac{L d}{20} + 1.51055$$

et dans le traçage

$$\underset{\text{traçage}}{E(L\eta)} = L 20 - 1.5 = 1.4958$$

La variance σ_i^2 est égale à la différence de ces deux valeurs moyennes multipliées par 2α

(34) $$\sigma_i^2 = 2\alpha\left[- \frac{L d}{20} \quad 0.0147 \right]$$

Ou, en reprenant le mètre comme unité

(35) $$\sigma_i^2 = \frac{2\alpha}{20} L \frac{10}{d} - \frac{2.95}{100}\alpha$$

soit, pour d = 10 cm

(36) $$\sigma_i^2 = \frac{46}{100}\alpha - \frac{2.95}{100}\alpha = \frac{43}{100}\alpha$$

On remarquera que le terme $\frac{2.95}{100}\alpha$ est presque négligeable. Le terme principal $\frac{1}{20}\, 2\alpha\, L\frac{10}{d}$ représente le quotient par le nombre 20 de sondages, de la variance $2\alpha L\frac{10}{d}$ d'un sondage dans sa "zone d'influence". Le terme correctif $\frac{2.95}{100}\alpha = \frac{0.58}{20}\alpha$ est égal au quotient par 20 du terme 0.58α qui représente l'élément préférentiel de la corrélation entre la " zone d'influence" et le sondage placé en son centre, terme peu différent de l'expression $\alpha/2$ qui apparait classiquement dans l'étude des mailles rigides.

Autrement dit, la variance de l'estimation d'un traçage par n sondages équidistants peut être considérée comme donnée par la formule classique des mailles rigides :

(37) $$\sigma_E^2 = \frac{\sigma^2 - \sigma_J^2 - \alpha/2}{n}$$

où σ^2 est la variance d'un sondage dans le traçage et σ_J^2 la variance de la zone d'influence d'un sondage dans le traçage. L'expression σ^2/q serait exagérément pessimiste (ici : $\frac{\sigma^2}{q} = \frac{2\alpha\, L\, 2000}{20} = \frac{76}{100}\alpha$ au lieu de $\frac{43}{100}\alpha$)

Estimation de S_2

On prend comme estimateur de la teneur y de S_2 l'expression habituelle :

$$v = A\,(y_1\, y_2 \cdots y_{100})^{\frac{1}{100}}$$

et la variance de cet estimateur est

$$\sigma_v^2 = \frac{\sigma_i^2}{5} + \alpha\frac{\pi}{2}\frac{S_2}{L^2} = \frac{8.6}{100}\alpha + \alpha\frac{\pi}{2}\frac{4.10^4}{10^6}$$

(38) $$\sigma_v^2 = \alpha\left[\frac{8.6}{100} + \frac{6.28}{100}\right] = \frac{14.88}{100}\alpha$$

C - Estimation du gisement S.

Cette estimation se fera à partir de l'estimateur u de S_1, formule (3?) et v de S_2, formule (38), soit :

(39) $$v = A\,u^{a}\,\sigma^{1-a}$$

Si l'on appliquait les formules (18) on aurait :

(40) $$\begin{cases} a = \dfrac{\sigma_\sigma^2 - \sigma_{u\sigma}}{\sigma_u^2 + \sigma_\sigma^2 - 2\sigma_{u\sigma}} \\[2ex] \sigma_v^2 = \dfrac{\sigma_u^2\,\sigma_\sigma^2 - \overline{\sigma_{u\sigma}}^2}{\sigma_u^2 + \sigma_\sigma^2 - 2\sigma_{u\sigma}} \end{cases}$$

Nous allons, en fait, constater numériquement que les formules (40) conduisent à des absurdités. Supposons, en effet, que nous connaissions non pas des estimations u et σ des teneurs de S_1 et S_2, mais leurs valeurs réelles x et y. A ce moment, la valeur réelle de la teneur du gisement est exactement :

$$z = \frac{S_1 x + S_2 y}{S_1 + S_2} = \frac{2x + y}{3}$$

....../...

et la variance de l'estimateur doit être nulle. Or les formules (40) conduisent à une valeur de a différente de $\frac{2}{3}$ et à une valeur non nulle de la variance.

En effet, les variances dans S des panneaux S_1 et S_2, calculées par les formules du rectangle sont :

$$(41)\quad \begin{cases} \sigma^2_{S_1} = 0.57948\,\alpha \\ \sigma^2_{S_2} = 1.29060\,\alpha \\ \sigma_{S_1 S_2} = -0.92713\,\alpha \end{cases}$$

avec

$$(42)\quad S_1^2\sigma^2_{S_1} + S_2^2\sigma^2_{S_2} + 2S_1S_2\,\sigma_{S_1S_2} = 0$$

On déduit des formules (40) a = 0.617, au lieu de 0.666, ce qui est manifestement absurde. La raison en est que les panneaux S_1 et S_2 ne peuvent en aucune façon être considérés comme implantés au hasard dans S. Dans un gisement fictif entourant S, les covariances σ_{SS_1} et σ_{SS_2} ne sont nullement égales, sauf si les panneaux S_1 et S_2 sont eux-mêmes égaux. Autrement dit, nous devons utiliser le système d'équations (10) et non le système (11).

Le système (10) s'écrit :

$$\begin{cases} a\,\sigma^2_u + (1-a)\,\sigma_{uv} - \sigma_{\zeta u} = \lambda \\ a\,\sigma_{uv} + (1-a)\,\sigma^2_v - \sigma_{\zeta v} = \lambda \end{cases}$$

soit, par différence :

$$(43)\quad a = \frac{\sigma^2_v - \sigma_{uv} + \sigma_{\zeta u} - \sigma_{\zeta v}}{\sigma^2_u + \sigma^2_v - 2\sigma_{uv}}$$

On peut admettre $\sigma_{\zeta u} = \sigma_{SS_1}$, $\sigma_{\zeta v} = \sigma_{SS_2}$, $\sigma_{uv} = \sigma_{S_1S_2}$ et écrire :

$$(44)\quad \begin{cases} \sigma_{\zeta u} = \sigma_{SS_1} = \dfrac{S_1\sigma^2_{S_1} + S_2\,\sigma_{S_1S_2}}{S_1+S_2} \\ \sigma_{\zeta v} = \sigma_{SS_2} = \dfrac{S_1\sigma_{S_1S_2} + S_2\,\sigma^2_{S_2}}{S_1+S_2} \end{cases}$$

1 - Si l'on connait les teneurs réelles x et y de S_1 et S_2 ($u = x$, $\sigma = y$)
la formule (43) donne :

$$a = \frac{\sigma^2_{S_2} - \sigma_{S_1 S_2} + \dfrac{S_1(\sigma^2_{S_1} - \sigma_{S_1 S_2}) - S_2(\sigma^2_{S_2} - \sigma_{S_1 S_2})}{S_1 + S_2}}{\sigma^2_{S_1} + \sigma^2_{S_2} - 2\sigma_{S_1 S_2}} = \frac{S_1}{S_1 + S_2}$$

et l'on vérifie par la formule (9) que la variance de z et x et y fixés est nulle. Nous échappons donc aux conséquences absurdes des formules (40)

2 - Pour le problème pratique qui nous occupe, nous ne connaissons pas les valeurs réelles de x et y, mais leurs estimations u et σ. Les valeurs numériques des variances doivent être prises dans S. Par exemple σ^2_u est égal à la somme du σ^2_u de la formule (29) et de la variance $\sigma^2_{S_1}$ de la formule (41) soit :

$$(45) \quad \begin{cases} \sigma^2_u = (0.57948 + 0.02578)\alpha = 0.60528\,\alpha \\ \sigma^2_\sigma = (1.39060 + 0.14888)\alpha = 1.53946\,\alpha \\ \sigma_{u\sigma} = -\,0.92713\,\alpha \\ \sigma_{zu} = +\,0.07727\,\alpha \\ \sigma_{z\sigma} = -\,0.15455\,\alpha \end{cases}$$

La formule (43) donne alors :

$$(46) \qquad \alpha = 0.675 \qquad\qquad 1 - \alpha = 0.325$$

x étant connu avec une meilleure précision que y, le poids du panneau S_1 est légèrement majoré vis à vis de la valeur normale 2/3 = cette majoration est cependant très faible, et on se rend bien compte qu'on ne perdrait pas beaucoup d'information en s'en tenant à la valeur régulière 2/3.

Enfin, la variance d'échantillonnage, variance liée à u et σ fixés est donnée par la formule (9). On trouve :

$$(47) \qquad \sigma^2_w = \frac{2.71}{100}\,\alpha$$

Remarque : Si l'on avait pris la valeur régulière $a = \frac{2}{3}$, la variance d'échantillonnage, donnée par (9) se réduirait à

$$\sigma^2_w = \left(\frac{2}{3}\right)^2 \frac{2.58}{100}\,\alpha + \left(\frac{1}{3}\right)^2 \frac{14.88}{100}\,\alpha = \frac{2.80}{100}\,\alpha$$

un petit peu plus grande que la valeur optimum (47)

ALGER - Février 1959