M. Matheron

Note statistique N° 6

Existence d'une erreur systématique d'échantillonnage.

Il ne s'agit pas ici d'une erreur imputable à l'imperfection du mode de prélèvement, par exemple de l'excès systématique de fines dans une pelletée; Krige a montré dans sa thèse (voir référence I, note ST.n° 4) que, même avec un mode d'échantillonnage parfaitement aléatoire, du seul fait que l'on prétend estimer la teneur d'un panneau de volume V_i au moyen d'un ou de plusieurs échantillons de volume global inférieur à V_i, l'on introduit une erreur systématique qui conduit à sous estimer la teneur du panneau, si elle est basse, et à la surestimer, si elle est élevée.

En effet, reprenant les notations de la Note n° 4, désignons par σ_μ^2 la variance des panneaux de volume V_i, par m la teneur moyenne du gisement, par m_i la teneur du panneau, par x l'évaluation de m_i fournie par le mode d'échantillonnage adopté, qui peut être quelconque, sous réserve que x suive une loi lognormale de moyenne m_i et de variance σ_i^2; soient enfin μ_i la médiane des x dans le panneau, et μ celle des x dans le gisement, μ étant également la médiane des μ_i.

On connait x et on cherche à estimer m_i. Déterminons donc la loi de probabilité suivie par m_i lorsque x est fixé. On a

$$m_i = \mu_i \, e^{\sigma_i^2/2}$$

Pour x fixé, le coefficient de corrélation de $L\mu_i$ et Lx étant σ_μ/σ [Note St 4, formule (2)], μ_i est lognormale et :

$$(1)\quad \begin{cases} \mathcal{E}(L\mu_i) = \sigma_i^2/\sigma^2 \, L\mu + \sigma_\mu^2/\sigma^2 \, Lx \\ D^2(L\mu_i) = \dfrac{\sigma_\mu^2 \sigma_i^2}{\sigma^2} \end{cases}$$

Par suite m_i est lognormale avec :

$$(2)\quad \begin{cases} \mathcal{E}(Lm_i) = \dfrac{\sigma_i^2}{\sigma^2} L\mu + \dfrac{\sigma_\mu^2}{\sigma^2} Lx + \dfrac{\sigma_i^2}{2} = \dfrac{\sigma_i^2}{\sigma^2} Lm + \dfrac{\sigma_\mu^2}{\sigma^2} Lx \\ D^2(Lm_i) = \dfrac{\sigma_\mu^2 \sigma_i^2}{\sigma^2} \end{cases}$$

Et la valeur moyenne de m_i a pour logarithme

$$L[\mathcal{E}(m_i)] = \sigma_i^2/\sigma^2 \, Lm + \sigma_\mu^2/\sigma^2 \, Lx + \sigma_i^2\sigma_\mu^2/2\sigma^2$$

et :

$$(3)\quad \boxed{\mathcal{E}(m_i) = m^{\frac{\sigma_i^2}{\sigma^2}} \, e^{\frac{\sigma_i^2 \sigma_\mu^2}{2\sigma^2}} \, x^{\frac{\sigma_\mu^2}{\sigma^2}}}$$

<u>C'est cette valeur qui doit être prise comme estimation de la teneur d'un panneau pour lequel un mode d'échantillonnage de variance σ_i^2 a donné la valeur x.</u> Ce n'est que pour $\sigma_i = 0$ que la valeur moyenne de m_i est égale à x

Pratiquement σ_i n'est jamais nul, et la teneur x doit être multipliée par le coefficient correcteur K (<u>Block Plan Factor de Krige</u>):

$$(4)\quad \boxed{K = e^{\frac{\sigma_i^2 \sigma_\mu^2}{2\sigma^2}} \left(\frac{x}{m}\right)^{-\frac{\sigma_i^2}{\sigma^2}}}$$

K est plus grand que I lorsque X est inférieur a $m\,e^{\sigma_\mu^2/2}$

<u>La valeur probable de la teneur d'un bloc de minerai pour lequel l'échantillon prélevé à une teneur x n'est pas x, mais Kx.</u>
Cela signifie, de façon précise, que, la teneur moyenne d'un grand nombre de blocs échantillonnés à la valeur x, n'est pas x, mais Kx.

<u>Le rôle du coefficient K est de corriger chaque mesure individuelle x.</u>
Mais si l'on subsistue à la population des x celle des Kx, on obtient une population lognormale de même valeur moyenne m, mais, cette fois, trop concentrée, puisque sa variance est $\frac{\sigma_\mu^4}{\sigma^2} < \sigma_\mu^2$

Il n'y a rien que de naturel, puisqu'en assimilant m_i à sa valeur probable Kx, on a négligé la variance de m_i pour x fixé, variance donnée par (2).

La somme de ces deux variances $\frac{\sigma_\mu^2 \sigma_i^2}{\sigma^2} + \frac{\sigma_\mu^4}{\sigma^2}$

est naturellement égale à la variance σ_μ^2 des blocs.

On peut dire que le coefficient K permet de se faire une meilleure idée de chaque portion du gisement.

Si au contraire on cherche une représentation d'ensemble, il faut avoir recours à la loi lognormale de moyenne M et de variance σ_μ^2

Pratiquements les résultats d'analyse des échantillons permettent de déterminer σ et m. A moins donc que le mode d'échantillonnage ne soit presque parfait (σ_i très petit), on ne connait que la teneur moyenne du gisement.

<u>Pour déterminer le coefficient correcteur K, aussi bien que pour fixer, la répartition du tonnage dans les tranches de teneurs, il est obligatoire de connaître la variance σ_i^2 du mode d'échantillonnage, donc de prélever sur un même bloc et d'analyser un nombre suffisant d'échantillons semblables.</u> Connaissant σ et σ_i on en déduit la variance réelle par différence,

$$\boxed{\sigma_\mu^2 = \sigma^2 - \sigma_i^2}$$