

REVUE DE STATISTIQUE APPLIQUÉE

R. DUVAL

Genèse de certaines distributions dissymétriques

Revue de statistique appliquée, tome 9, n° 1 (1961), p. 75-88

http://www.numdam.org/item?id=RSA_1961__9_1_75_0

© Société française de statistique, 1961, 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/>

GENÈSE DE CERTAINES DISTRIBUTIONS DISSYMÉTRIQUES

par R. DUVAL

Ancien Élève de l'École polytechnique

Ingénieur civil des mines

On constate, lorsqu'on échantillonne un gisement de minerai, que la distribution des teneurs de ce gisement est généralement dissymétrique, le mode de la courbe représentant la distribution est reporté du côté des teneurs basses. Celles-ci naturellement ne peuvent jamais être négatives.

Il est d'ailleurs nécessaire d'ouvrir d'abord une parenthèse au sujet de cette expression de distribution des teneurs d'un gisement. On entend par là la distribution des teneurs moyennes de chacune des unités d'échantillonnage en lesquelles on subdivise par la pensée l'ensemble du volume du gisement qui est rempli d'une matière continue. Il y a donc en fait autant de distributions qu'il y a de conceptions d'unités d'échantillonnage, c'est-à-dire une infinité, mais ces distributions doivent avoir la même moyenne. On ne pourrait donc connaître cette distribution qu'après avoir vidé complètement le gisement.

En fait, on estime cette distribution d'après l'échantillonnage du gisement qui se fait généralement par la technique de l'échantillonnage systématique, en implantant par exemple des sondages au diamant aux sommets d'un quadrillage appliqué sur la carte du gisement, et en déterminant la teneur de chacun des prélèvements élémentaires (dits carottes) ainsi effectués. C'est sur cette sous-population de déterminations qu'on constate très souvent la dissymétrie de l'histogramme des fréquences dont il a été parlé au début. La raison de cette dissymétrie a fait l'objet de différentes hypothèses dont la critique va être ci-dessous.

EXPLICATIONS DONNEES DE LA DISSYMETRIE CONSTATEE

Divers auteurs ont voulu voir dans ces distributions dissymétriques, les uns la manifestation d'une loi binomiale, les autres d'une loi lognormale.

LOI BINOMIALE

Les défenseurs de la loi binomiale s'appuient sur le fait indubitable qu'en fait un gisement se compose élémentairement de particules : les unes de minéral pur (galène par exemple pour le plomb, c'est-à-dire des cristaux de sulfure de plomb), les autres de gangue stérile. Ils assimilent alors l'opération d'échantillonnage à un tirage au hasard d'un certain nombre n de particules dans une population très nombreuse où la proportion des particules de minéral est p . Dans ces conditions, le nombre de particules de minéral rencontrées dans chaque prélèvement de n particules, donc, disent-ils, la teneur doit obéir à une loi binomiale. On peut d'abord objecter qu'au-

cune des conditions requises pour obtenir une distribution binomiale n'est remplie dans ce schéma. Les particules ne sont pas identiques en forme, poids, volume, elles ne sont pas initialement parfaitement mélangées, leur nombre prélevé n n'est certainement pas constant, de plus elles sont agglomérées entre elles, etc.

Mais admettons néanmoins provisoirement que ces conditions soient remplies, chacun des prélèvements élémentaires effectués se composerait, en raison de la grandeur habituelle des parcelles de minéraux, d'un nombre élevé n de particules. Or on sait que, dans ces conditions, une loi binomiale tend rapidement vers une loi normale. On ne devrait donc pas alors trouver de distributions dissymétriques.

Il s'élève encore une autre objection. Nous avons vu ci-dessus qu'en fait, d'après la conception de base de cette hypothèse, c'est le nombre de particules de minéral ou leur proportion dans le prélèvement élémentaire qui devrait obéir à une loi binomiale.

Si donc on appelle n_1 le nombre de particules de minéral et n_2 le nombre de particules de gangue avec $n_1 + n_2 = n$ c'est la variable

$$x = \frac{n_1}{n_1 + n_2}$$

qui devrait obéir, d'après cette théorie, à la loi binomiale. Mais en fait on prélève un volume et la teneur dans ce volume est, en appelant d_1 la densité du minéral et d_2 la densité de la gangue :

$$z = \frac{n_1 d_1}{n_1 d_1 + n_2 d_2} = \frac{n_1}{n_1 + n_2 \frac{d_2}{d_1}}$$

Lorsque d_1 s'écarte très sensiblement de d_2 comme c'est le cas fréquent, les deux variables x et z sont différentes, elles sont liées par la relation

$$z = \frac{x}{x + (1 - x) \frac{d_2}{d_1}}$$

Si donc x obéissait à une loi binomiale, z , la teneur, suivrait une loi différente.

LOI LOGNORMALE

Les partisans de la loi lognormale sont très absolus, car ils considèrent qu'elle est la manifestation de l'application de la loi générale d'action de masse ou d'effet proportionnel tant aux teneurs de gisement proprement dites qu'aux "accumulations" c'est-à-dire aux produits des teneurs par l'épaisseur du gisement correspondante, et ils soutiennent qu'en conséquence la loi lognormale s'applique d'une façon complètement générale. Nous ne connaissons pas, en ce qui nous concerne, de théorie solidement construite montrant comment on déduit une loi de distribution lognormale des teneurs et des accumulations de ces lois d'action de masse ou d'effet proportionnel; ce que nous connaissons c'est seulement une simple affirmation, ou un acte de foi si l'on préfère, et une tentative faite pour démontrer, au prix d'hypothèses audacieuses et qui, croyons-nous, n'ont pas fait l'objet de vérifi-

cations, la permanence de la lognormalité, quelle que soit l'unité d'échantillonnage considérée; cette démonstration était en effet nécessaire pour pouvoir soutenir la généralité de cette loi lognormale.

Quelle que soit cette théorie, si elle existe, nous serions surpris qu'elle échappe à l'objection présentée ci-dessus à la loi binomiale au sujet de la différence des distributions du x et du Z ; en tout cas s'il est indubitable que, dans d'assez nombreux cas, une loi lognormale compense convenablement des distributions constatées expérimentalement de teneurs ou d'accumulations, des vérifications faites sur quelques gisements (référence 1) ne confirment pas qu'elle s'applique d'une manière générale.

D'ailleurs l'application d'une généralité simultanée pour les teneurs et les accumulations soulève une autre objection. Comme le logarithme d'un produit est la somme des logarithmes des facteurs et que la somme ou la différence de 2 variables indépendantes obéissant à la loi normale est une variable également normale, il en résulte que si la teneur et l'accumulation obéissent à une loi lognormale, l'épaisseur du gisement doit également obéir à une loi lognormale. Comme généralement, au point de vue géologique, l'épaisseur du gisement dépend de circonstances tectoniques ou stratigraphiques qui sont antérieures à l'arrivée de la minéralisation, on ne voit pas bien ce qu'une loi d'action de masse ou d'effet proportionnel aurait à voir dans cette affaire.

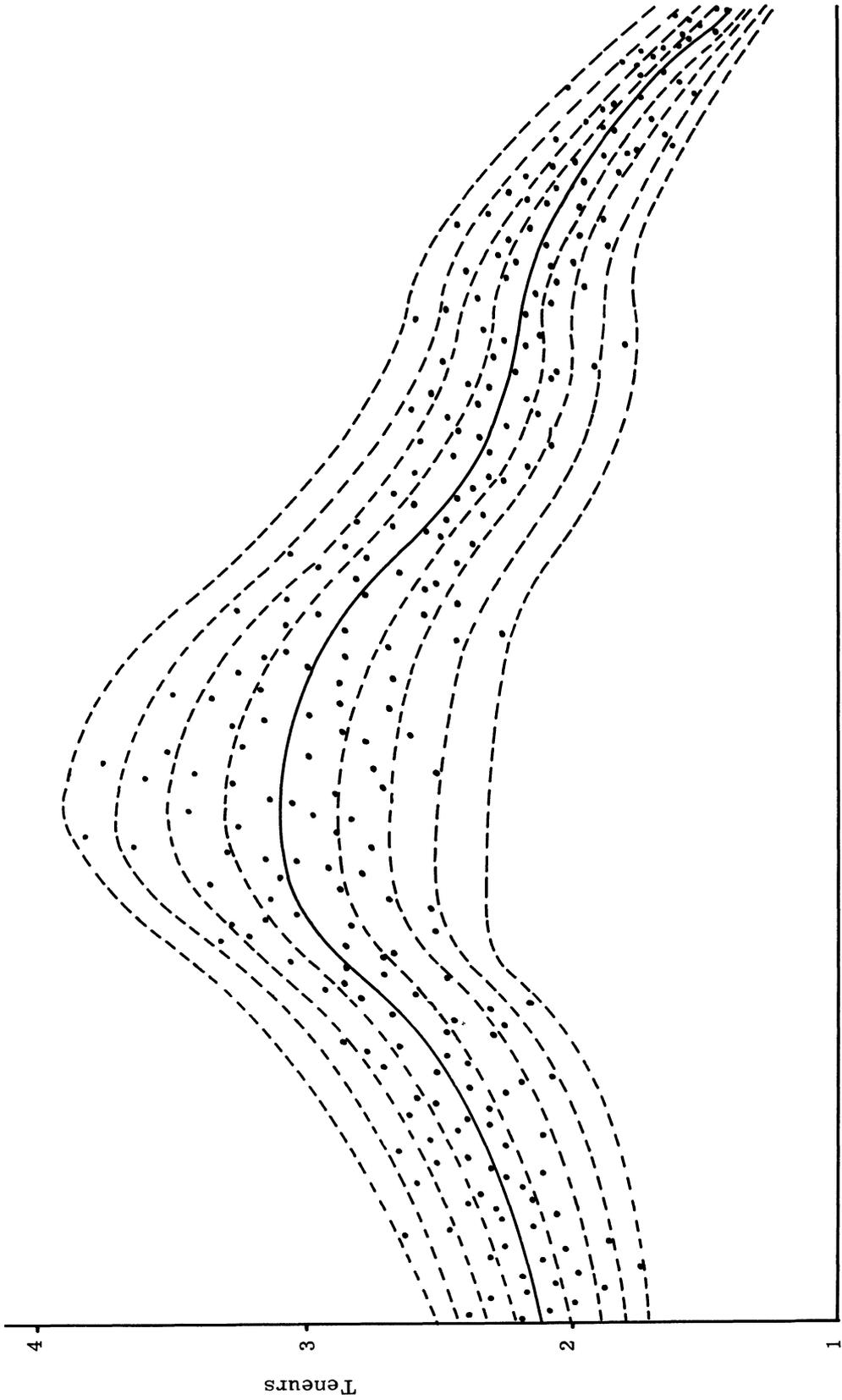
Cette question de la généralité de la loi lognormale a pratiquement de l'importance car ses défenseurs préconisent pour l'évaluation de la teneur moyenne du gisement l'emploi d'un certain estimateur qui donne des résultats erronés si la distribution n'est pas effectivement lognormale (référence 2).

Or on se contente assez souvent de peu pour vérifier qu'une distribution est lognormale, car beaucoup d'auteurs font seulement cette vérification en employant la technique de la droite de Henry qui peut être assez subjective et parfois aboutir à des résultats erronés comme nous avons eu l'occasion de le constater.

Les objections exposées ci-dessus paraissent donc rendre tout au moins douteux le bien fondé des explications étudiées ci-dessus. C'est donc une raison pour en chercher une autre.

NOUVELLE EXPLICATION SUGGEREE

L'idée de cette explication nous est venue en examinant un modèle de distribution de gisement établi en s'inspirant de la distribution des teneurs d'un gisement de plomb des Cévennes et qui est représenté sur la figure 1. Ce modèle, établi en vue d'une étude sur des modalités d'échantillonnage des gisements, contient une courbe en trait plein qui représente, en fonction des abscisses qui sont les distances sur le terrain, la teneur moyenne du gisement au point du terrain considéré. Les points marqués représentent en ordonnées la teneur effectivement constatée au point marqué en abscisses; ils se dispersent en ordonnées autour de la courbe moyenne suivant une loi qui a été admise ici être une loi normale, mais c'est uniquement pour éviter l'arbitraire, la loi peut être différente. La dispersion autour de la courbe moyenne augmente en même temps que la teneur moyenne. Ces conditions n'ont pas été fixées arbitrairement, elles correspondent à celles rencontrées souvent dans les gisements plombo-zingueux notamment. Comme ce genre de graphique sera encore utilisé plus loin, ajoutons pour préciser sa construction qu'il n'y a jamais dans chaque colonne de 1 mm du papier quadrillé



Distances

Figure 1

qu'un seul point représentatif et qu'afin d'assurer une distribution en ordonnées sensiblement normale, les points ont été ici respectivement répartis au sentiment proportionnellement aux nombres 25, 16, 7, 2 dans chacune des bandes marquées par des traits interrompus de chaque côté de la courbe moyenne et dont la largeur de chacune représente un écart médian. En abscisses on s'est efforcé de conserver une distribution sensiblement uniforme dans chaque bande. On peut considérer, dans le cas présent, cette distribution comme celle des teneurs des carottes extraites d'une série de sondages jointifs le long d'un alignement sur le terrain.

L'examen de ce graphique indique tout de suite qu'en raison de l'existence d'un maximum de la courbe moyenne et de l'accroissement de la dispersion avec l'augmentation de la teneur moyenne le nombre des points décomptés dans une bande horizontale de 1 cm de largeur par exemple est dans la partie haute du dessin inférieure au nombre qu'on rencontre dans la partie inférieure. L'idée vient donc d'établir un histogramme en fréquence de la distribution des teneurs en décomptant dans chaque bande horizontale correspondant à 1 cm d'ordonnées (teneurs) le nombre de points rencontré.

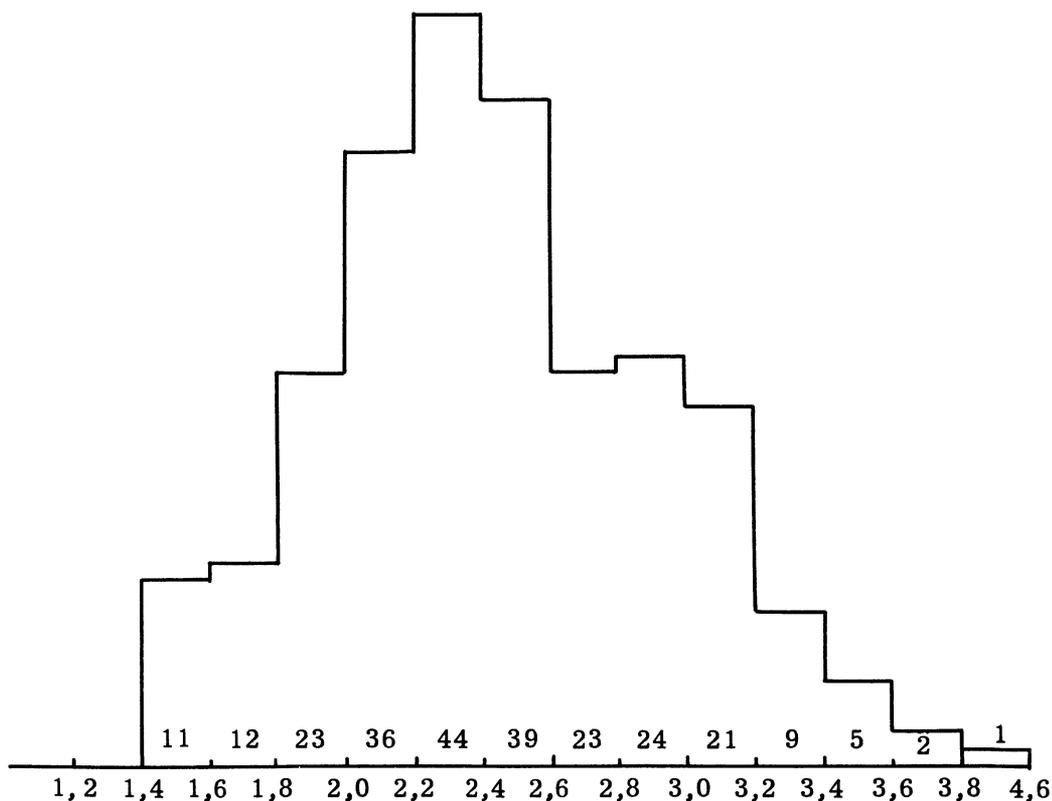


Figure 2

La figure 2 représente l'histogramme ainsi établi correspondant à la figure 1. On constate qu'il paraît bien en effet correspondre à une distribution dissymétrique avec mode reporté du côté des teneurs basses.

Nous avons donc étudié de plus près cette question, afin de déterminer les conditions auxquelles devait satisfaire une distribution de teneurs de gisement, constituée suivant ces principes, pour avoir un histogramme plus nettement dissymétrique comme on en rencontre souvent.

ETUDE STATISTIQUE DE LA GENESE DES DISTRIBUTIONS DISSYMETRIQUES DE TENEURS

Modèles simplifiés

1/ Nous avons étudié à cet effet d'abord un modèle schématique (figure 3) établi suivant les principes généraux indiqués ci-dessus, dans lequel la courbe des moyennes est toutefois remplacée par une ligne droite et le maximum

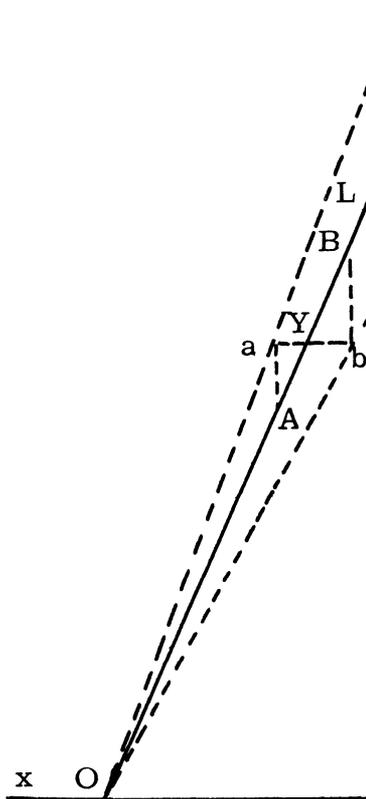


Figure 3

par un angle aigu réduit à un seul versant (la figure complète étant admise être symétrique). Les abscisses et les ordonnées ont la même signification que dans la figure 1.

Nous avons recherché par le calcul quel était le tracé de la courbe de la densité de probabilité en fonction de la teneur quand le point moyen de la distribution se déplace du point O au point L sur la droite en trait plein en admettant pour faciliter l'intégration qu'on avait affaire à une courbe de distribution en toit (figure 4) dans le sens des ordonnées au lieu d'avoir la distribution normale de la figure 1.

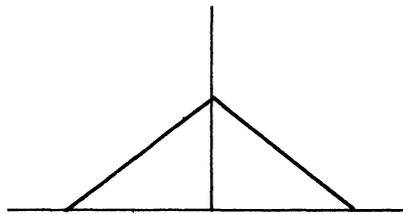


Figure 4

Les 2 droites en trait interrompu de la figure 3 encadrant la droite OL marquent les limites de l'amplitude de la dispersion de la distribution en toit. On voit donc que cette amplitude est proportionnelle à la teneur moyenne (coefficient de dispersion constant).

Dans ces conditions, si l'on considère une bande horizontale de largeur dy placée par exemple en ab sur la figure, toutes les distributions qui peuvent envoyer en direction verticale des points représentatifs dans cette bande sont évidemment seulement celles dont les points moyens sont compris entre les points A et B et on peut, pour chaque point de AB, déterminer la probabilité pour ledit point représentatif de tomber dans le segment dy ; cette probabilité sera fonction de x . En intégrant de a à b seulement on aura donc la densité de probabilité à la cote y . Cette intégration ne présente pas de difficulté et on trouve une densité de probabilité indépendante de la cote y et dépendant seulement des angles que font avec l'horizontale les droites inclinées du graphique. Toutefois la densité de probabilité baisse progressivement jusqu'à 0 à partir du moment où le point B arrive en L, car le segment, ab , s'ampute progressivement sur sa droite jusqu'à disparaître totalement lorsque le point A arrive en L. On obtient donc une courbe de densité de probabilité du type de la figure 5.

Cette constance de la densité de probabilité sur une grande étendue, qui peut surprendre au premier abord, s'explique facilement lorsqu'on remarque que dans la figure 3 tout est homothétique par rapport au point O, et comme dans l'intégration les points n'ont pas de dimensions on se trouve en fait, quelle que soit la position du point y , en présence d'une figure identique. Ce raisonnement est d'ailleurs encore valable si on remplace la distribution en toit par exemple par une distribution normale à coefficient de dispersion constant.

Mais on constate sur la figure 5 l'absence d'un maximum de densité de probabilité qui évidemment ne correspond pas à ce qu'on observe dans les gisements. Ce modèle ne convient donc pas à l'objet que nous nous proposons.

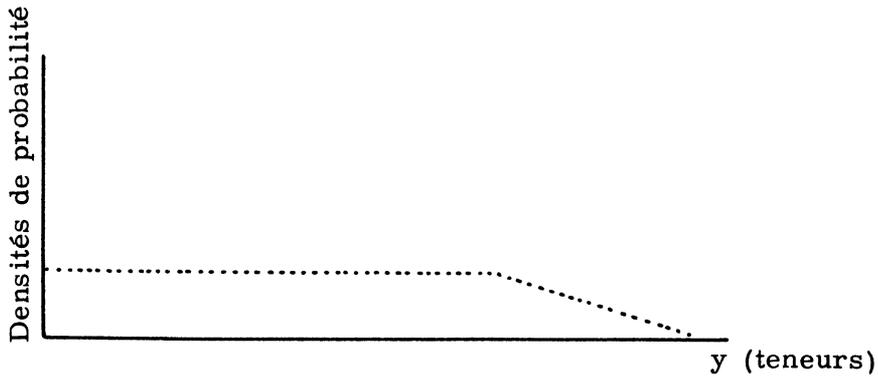


Figure 5

En effet dans notre cas les points représentatifs ont une dimension finie qui correspondrait par exemple au diamètre des carottes dont il a été parlé plus haut (1 seul point par colonne d'1 mm de largeur), il s'agit donc de distributions discrètes et non pas continues, l'emploi du calcul intégral n'est pas justifié, il faut faire des sommations point par point, et lorsque le point considéré de la droite OL se rapprochera du point O on ne pourra plus inscrire qu'un nombre limité de points au lieu d'une infinité et la densité de probabilité devra baisser jusqu'à O lorsqu'on se rapproche du point O au lieu de descendre suivant une verticale comme dans la figure 5.

2/ Nous avons alors vérifié sur un nouveau modèle (figure 6) analogue à celui de la figure 3, le bien fondé des considérations ci-dessus; en calculant d'une part par une sommation discrète, et en admettant une distribution normale, la courbe de la densité de probabilité et en vérifiant d'autre part par l'établissement d'une distribution expérimentale suivant les principes exposés à propos de la figure 1 que l'histogramme des fréquences de cette distribution expérimentale s'ajuste convenablement sur la courbe de densité de probabilité calculée. La seule différence avec la figure 1 est que pour la distribution expérimentale nous avons adopté pour la facilité du calcul, cette fois-ci, de part et d'autre de la droite, lieu géométrique des moyennes, 3 bandes de 1 écart-type chacune et limité l'amplitude de la distribution expérimentale à ces 3 écarts-types. La répartition des points dans ces 3 bandes a donc été faite, cette fois-ci, proportionnellement aux nombres 34, 14 et 2. Le calcul point par point de la courbe de densité de probabilité est naturellement long et fastidieux et nous avons limité ce calcul à celui d'un certain nombre de points critiques qui sont marqués d'une croix sur le graphique, ils permettent de définir convenablement l'allure de la courbe.

Nous reproduisons ci-dessous à titre d'exemple le calcul du point correspondant à $y = 18$ mm (cote 1,8).

En examinant la figure 6 on constate qu'un déplacement sur la droite OL correspondant à un déplacement en x de $\Delta x = \pm 1$ mm correspond d'autre part à un déplacement $\Delta y = \pm 3,6$ mm en ordonnée et à une variation de l'écart-type de $\Delta \sigma = \pm 0,4$ mm.

La probabilité qu'a un point représentatif de la distribution considérée de se trouver dans une bande horizontale de 1 mm de hauteur encadrant le point $y = 18$ mm, c'est-à-dire dans une bande comprise entre $y = 17,5$ mm et $y = 18,5$ mm, va être calculée. A cet effet, il va être examiné successivement dans chacune des bandes verticales de 1 mm de largeur (1 seul point

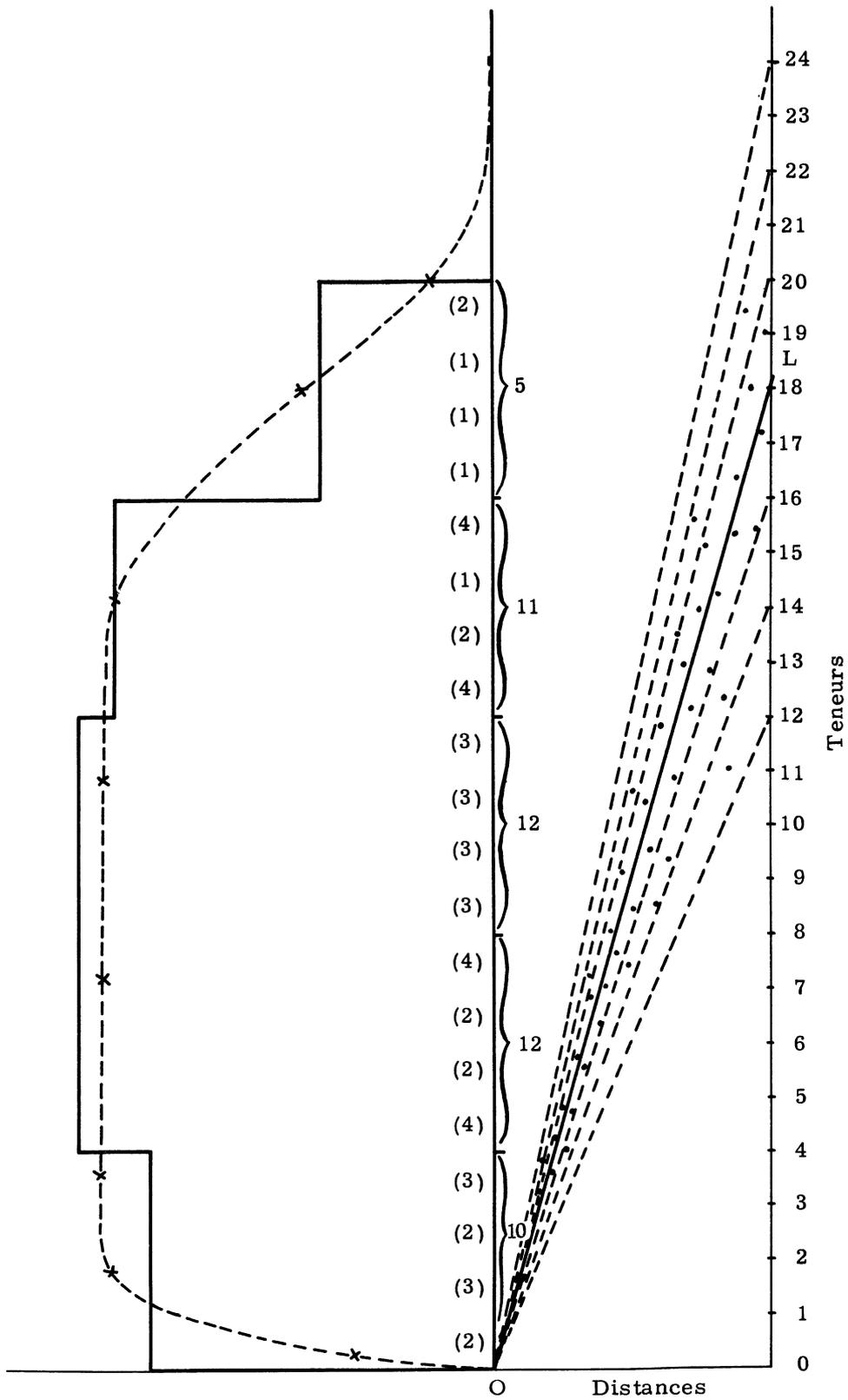


Figure 6

par bande verticale de 1 mm d'après la construction du graphique) les distributions normales ayant leur centre sur la droite OL et l'écart-type correspondant qu'indique la figure 6. On calcule alors, conformément au tableau 1, pour chacune de ces distributions au moyen d'une table de la fonction de Laplace-Gauss la probabilité pour que le point se place dans la bande horizontale définie ci-dessus comprise entre $y_1 = 17,5$ et $y_2 = 18,5$. Ce calcul a été poursuivi jusqu'à ce que les probabilités trouvées soient inférieures à 0,00001.

Tableau I
($y_1 = 17,5$, $y_2 = 18,5$, $\Delta y = \pm 3,6$, $\Delta \sigma = \pm 0,4$)

Paramètres des distributions		Ecart		Ecart réduits		Probabilité de la table		Probabilité dans la bande
y	σ	$\Delta y_1 = y_1 - y$	$\Delta y_2 = y_2 - y$	$\frac{\Delta y_1}{\sigma} = t_1$	$\frac{\Delta y_2}{\sigma} = t_2$	$\theta(t_1)$	$\theta(t_2)$	$\theta(t_1) + \theta(t_2)$
18	2	0,5	0,5	0,25	0,25	0,09871	0,09871	0,19742
y	σ	Δy_1	Δy_2	$\frac{\Delta y_1}{\sigma} = t_1$	$\frac{\Delta y_2}{\sigma} = t_2$	$\theta(t_1)$	$\theta(t_2)$	$\theta(t_2) - \theta(t_1)$
Points de OL avec $y < 18$								
14,4	1,6	3,1	4,1	1,937	2,652	0,47363	0,49480	0,02117
10,8	1,2	6,7	7,7	5,6...	6,4...	0,49999..	0,49999..	0,00000
Points de OL avec $y > 18$								
21,6	2,4	4,1	3,1	1,708	1,291	0,45618	0,40164	0,05454
25,2	2,8	7,7	6,7	2,750	2,392	0,49702	0,49162	0,00540
28,8	3,2	11,3	10,3	3,531	3,218	0,49979	0,49936	0,00043
32,4	3,6	14,9	13,9	4,138	3,801	0,49998	0,49994	0,00004
							Total général : 0,2790	

La densité de probabilité dans la bande horizontale de 1 mm de largeur à l'ordonnée 18 est donc de 0,2790.

La courbe de densité de probabilité de la figure 6 définie par le calcul des points marqués d'une croix a été tracée sur la figure 6 en trait interrompu en fonction de y (la teneur), en adoptant pour la densité de probabilité une échelle arbitraire.

La distribution expérimentale des points représentatifs de la figure 6 a été tracée suivant les principes déjà exposés plus haut. Les nombres de points décomptés dans chaque bande horizontale de 10 mm sont ceux inscrits entre parenthèse sur la figure 6. On constate, par exemple, que la bande horizontale de 10 mm de largeur dans laquelle se trouve ce point $y = 18$ mm contient dans la distribution expérimentale 3 points; en se guidant sur la courbe de densité de probabilité pour déterminer la probabilité qu'a un point de se trouver dans chaque bande de 1 mm contenue dans ces 110 mm, on trouve un total de 2,603 points ce qui est convenablement concordant.

Pour construire l'histogramme des fréquences du graphique (courbe en escalier en trait plein), on a adopté comme base uniforme des rectangles une longueur de 40 mm afin d'aplanir les variations locales et pour déterminer l'échelle des fréquences de l'histogramme on a fait coïncider le côté opposé à la base du rectangle correspondant avec le point $y = 144$ mm de la courbe de densité de probabilité. Les autres côtés opposés à la base des différents rectangles ont été déterminés en fonction de l'échelle ainsi déterminée et du nombre de points décomptés dans la bande horizontale correspondante de 40 mm de largeur. On constate un ajustement fort satisfaisant de l'histo-

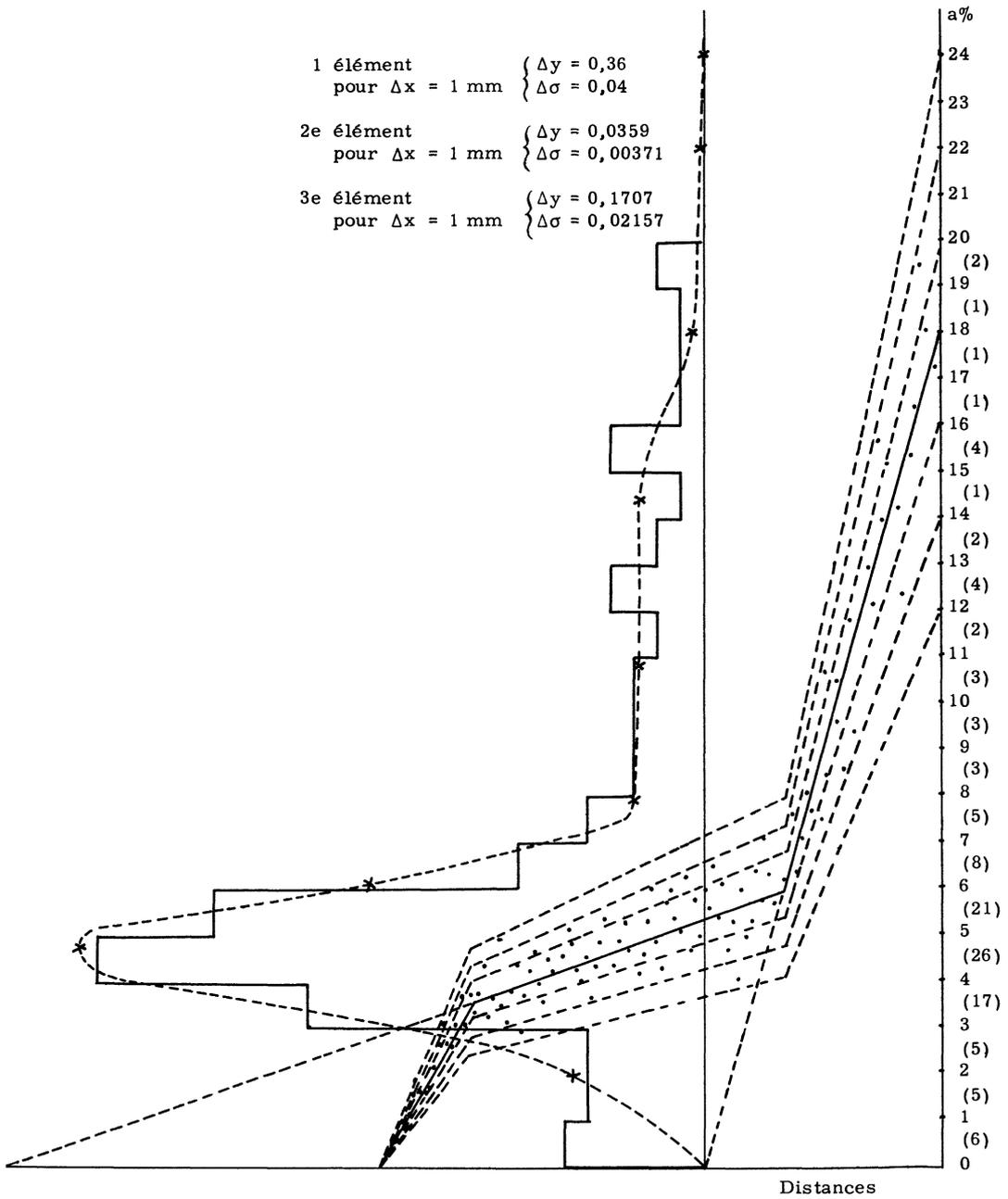


Figure 7

gramme sur la courbe de densité de probabilité calculée. On voit donc que les prévisions faites sont confirmées, la courbe de densité de probabilité présente bien l'aspect attendu; densité de probabilité tendant progressivement vers 0 à la partie supérieure, région à densité constante puis décroissance de nouveau zéro en se rapprochant du point 0.

On obtient donc du fait de l'existence du maximum de la courbe des teneurs moyennes et de l'augmentation de la dispersion avec la croissance de la teneur moyenne une disposition dissymétrique de la distribution des teneurs, et du fait de la distribution discrète des points représentatifs une chute nouvelle de la densité de probabilité au voisinage du point 0. Il y a donc maintenant existence d'un maximum de la densité de probabilité, mais le maximum présente un long palier, qui ne correspond pas à ce que l'on constate pour les distributions de teneur de gisement. Ceci est dû à la rectitude de la ligne des moyennes.

Il y a donc lieu d'examiner comment doit se modeler cette ligne pour qu'on puisse reproduire les distributions constatées dans la nature.

3/ A cet effet nous avons repris sur un nouveau modèle représenté par la figure 7 le calcul de la courbe de densité de probabilité et l'établissement de la distribution expérimentale d'après les principes employés pour la figure 6. Ce modèle a cette fois-ci une ligne des teneurs moyennes constituée par une ligne brisée à trois pentes. L'histogramme, qui a été tracé ici par bandes horizontales de 1 cm, s'adapte encore fort bien sur la courbe de densité de probabilité calculée. On obtient maintenant un maximum marqué.

Cette étude permet d'arriver aux conclusions suivantes concernant le modelé recherché de la courbe des teneurs moyennes.

CONCLUSIONS

Si l'on a une courbe symétrique des teneurs moyennes à 5 courbures ayant l'allure de la figure 8a avec une dispersion des teneurs autour de cette courbe augmentant lorsqu'augmente la teneur moyenne on obtient une courbe des fréquences en fonction de la teneur ayant l'allure de la figure 8b.

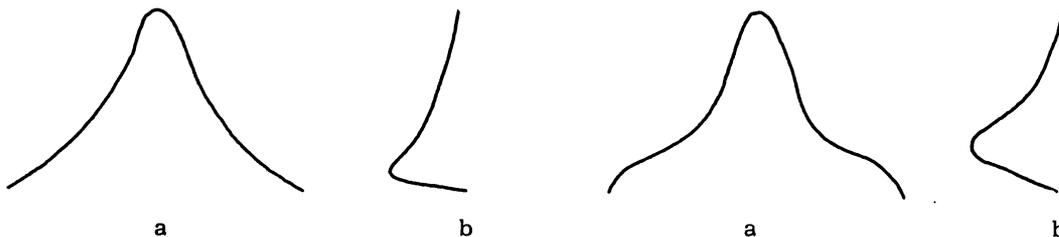


Figure 9

Figure 8

Si la courbe des teneurs moyennes a l'allure à 3 courbures seulement de la figure 9a on obtiendra une courbe de fréquences à chute plus brusque vers la teneur zéro ayant l'allure de la figure 9b.

Les figures 8b et 9b correspondent bien à des distributions rencontrées souvent dans la nature.

On peut vérifier ces conclusions sur un exemple expérimental de distribution. Voir figure 10.

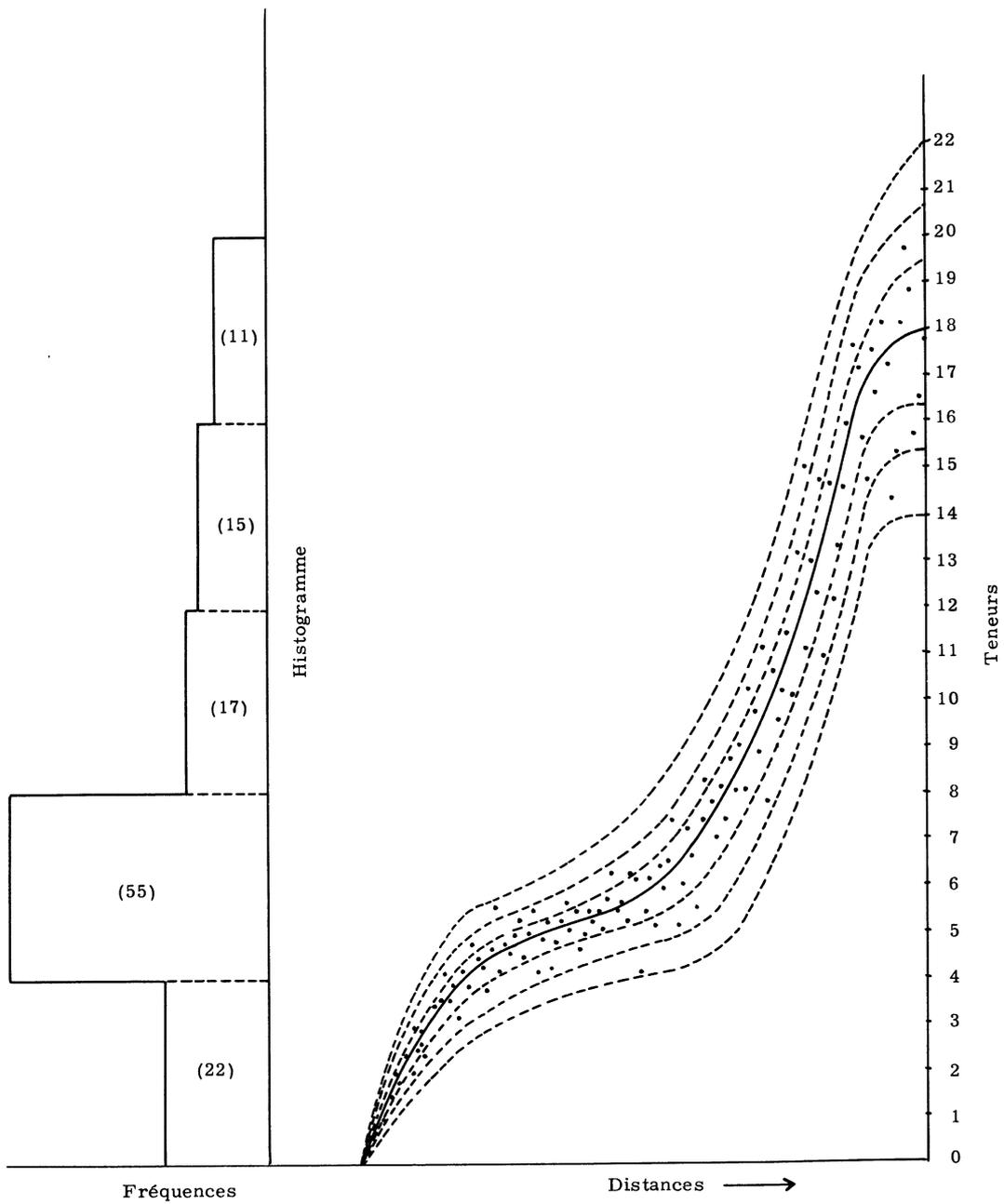


Figure 10

Quels peuvent être les phénomènes naturels donnant lieu à ces allures de courbes de teneurs moyennes et à ces dispersions augmentant avec les teneurs moyennes ? Pour répondre d'une manière irréfutable à ces questions, il faudrait d'abord savoir comment s'opère le phénomène de minéralisation lui-même, ce qui est encore du domaine des hypothèses, bien que les phénomènes des "Suffioni" de Toscane et des geysers puissent en donner quelque idée, tout au moins pour les minéralisations qui ne s'expliquent pas par une simple précipitation chimique. Dès maintenant, les prospecteurs considèrent généralement le maximum de teneur moyenne constaté comme ayant des chances de correspondre au point où l'injection minéralisante s'est produite avec toute l'énergie qu'elle contenait dans un milieu préexistant plus ou moins poreux et hétérogène où elle s'est répandue par diffusion ou écoulement. Une dispersion diminuant avec une teneur moyenne qui s'abaisse au fur et à mesure que l'épanchement s'étend dans un plus grand volume et que l'énergie interne s'amortit ne paraît pas contraire à la logique. Ces conditions suffisent déjà à créer un histogramme dissymétrique. Quant aux formes des courbes de teneurs moyennes vues dans les figures 8a et 9b, il appartient aux spécialistes des phénomènes de diffusion et d'écoulement de dire si elles ne peuvent pas s'expliquer aussi par ces phénomènes mêmes.

REFERENCES

- (1) R. DUVAL - "Contrôle de qualité et analyse statistique en matière de laveries métalliques". Annales des Mines Nov. 1957, p. 740-41 et suivantes.
- (2) R. DUVAL, R. LEVY et G. MATHERON - "Travaux de M. D.G. KRIGE sur l'évaluation des gisements dans les mines d'or de l'Afrique du Sud". Annales des Mines, Décembre 1955, p. 25-27.