

# REVUE DE STATISTIQUE APPLIQUÉE

JACQUES AGARD

**Mélange de deux populations normales et étude de quelques fonctions  $f(x, y)$  de variables normales  $x, y$**

*Revue de statistique appliquée*, tome 9, n° 4 (1961), p. 53-70

[http://www.numdam.org/item?id=RSA\\_1961\\_\\_9\\_4\\_53\\_0](http://www.numdam.org/item?id=RSA_1961__9_4_53_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/>

# MÉLANGE DE DEUX POPULATIONS NORMALES ET ÉTUDE DE QUELQUES FONCTIONS $f(x, y)$ DE VARIABLES NORMALES $x, y$ .

Jacques AGARD

Docteur ès-sciences

*Cet article, partiellement extrait de notre thèse<sup>(1)</sup> expose les méthodes utilisées pour étudier une population résultant du mélange dans des proportions inconnues d'observations issues de deux populations de variables normales  $(x, y)$  de même moyenne. Nous calculons en outre les lois des variables  $xy$ ,  $\frac{x}{y}$ , et  $\frac{x - y}{x + y}$  lorsque  $x$  et  $y$  sont deux variables normales corrélées de même moyenne et de même variance. Ces considérations théoriques sont utiles en particulier pour analyser la loi de répartition des coefficients de réflexion existant dans les profondeurs du sol qui sont à l'origine des réflexions observées sur les sismogrammes enregistrés lors de la prospection sismique.*

## I - ETUDE DES COEFFICIENTS DE REFLEXION -

La terre est un milieu solide dont la constitution varie avec la profondeur. On peut schématiquement représenter le sol comme la superposition de couches géologiques successives dans lesquelles la vitesse de propagation du son varie. Les discontinuités des constantes électriques ou des densités entraînent des réflexions, réfractions ou diffractions des ondes émises en surface, analogues à celles subies par un rayon lumineux lors de la traversée de milieux liquides. En simplifiant et en négligeant les variations de densité du sol, on peut dire qu'à la surface de séparation de deux couches géologiques correspond un coefficient de réflexion. Ce coefficient de réflexion mesuré à la profondeur  $h$  est :

$$k(h) = \frac{V(h) - V(h + dh)}{V(h) + V(h + dh)}$$

$V(h)$  et  $V(h + dh)$  étant les vitesses de propagation du son dans les deux couches voisines.

Pour connaître la constitution géologique du sol sans effectuer de forage, on procède à une explosion, et on étudie les retours d'énergie enregistrés par des sismographes à la suite des réflexions de l'onde émise par l'explosion, sur la surface des diverses couches géologiques.

L'analyse statistique des sismogrammes nous a conduit à étudier préalablement la loi de répartition des amplitudes des coefficients de réflexion.

On dispose dans certains cas d'échantillonnages de ces coefficients à la suite de forages au cours desquels on mesure la vitesse de propagation du son  $V(h)$  en fonction de la profondeur. L'ajustement des observations  $k(h)$  d'un sondage donné à une loi normale s'est révélé peu satisfaisant et il a semblé que les  $k(h)$  observés pouvaient provenir de la combinaison de deux populations normales de même moyenne et de variances différentes dans des proportions inconnues a priori. Ce fait nous a amené à étudier le problème de la discrimination des paramètres des deux lois, dans une première partie.

Par ailleurs, l'assimilation des vitesses de propagation  $V(h)$  observées à une loi normale étant satisfaisante, et les couples  $V(h)$ ,  $V(h + dh)$  ayant une corrélation approximativement stationnaire  $r$ , nous avons étudié quelle loi en résultait pour  $k(h)$  et accessoirement pour des fonctions plus simples  $\frac{V(h)}{V(h + dh)}$ ,  $V(h).V(h + dh)$  ... Ceci fait l'objet de la seconde partie de cet article.

## II - COMPOSITION DE DEUX VARIABLES NORMALES -

Les coefficients de réflexion  $k(h)$  sont calculés à partir des vitesses  $V(h)$  mesurées régulièrement tous les  $dh$  entre 0 et  $H$ , profondeur maximum. Leur ajustement à la loi normale a été tenté sur de nombreux exemples. La figure 1 fournit un exemple d'ajustement par la droite de Henri. On constate que l'ajustement n'est pas satisfaisant puisque nous avons trois alignements au lieu d'un seul. Ce phénomène a été observé dans les nombreux exemples étudiés. D'autres lois statistiques ne permettent pas un meilleur ajustement. L'existence des trois alignements sur la droite de Henri nous a paru provenir vraisemblablement du mélange des deux populations. Ces deux populations peuvent résulter du mélange dans les observations des coefficients de réflexion de deux couches lithographiques différentes pour lesquelles les variances des  $k$  diffèrent. Elles peuvent aussi s'expliquer par le mélange de coefficients de réflexion correspondant à des changements de couches géologiques importantes et de coefficients de réflexion plus faibles existant à l'intérieur des couches principales, ou enfin à l'addition aux coefficients de réflexion réels de faux coefficients de réflexion entraînés par les erreurs de mesure des vitesses : ces erreurs introduiraient de faibles variations de vitesse  $V(h)$  fictives.

Dans ces diverses hypothèses, la combinaison de deux lois normales, de moyenne nulle et d'écart-types différents, dans des proportions données fournit trois alignements sur un graphique de Henri. Les deux alignements extrêmes, pratiquement parallèles, correspondent à la population des variables normales à forte dispersion. L'alignement du centre correspond au mélange des deux populations. L'examen du graphique de Henri permet de se faire une idée approchée des paramètres. Il semble qu'à moins de disposer d'un très grand nombre d'observations, il soit difficile de déterminer les paramètres inconnus des deux populations avec précision. On dispose cependant d'une méthode mathématique d'estimation :

Appelons  $f(x)$  la densité de probabilité des coefficients de réflexion et soit :

- $p_1$  = proportion de la population 1,
- $p_2 = 1 - p_1$  = proportion de la population 2,
- $\sigma_1^2$  = dispersion de la population 1,
- $\sigma_2^2$  = dispersion de la population 2,

les moyennes sont nulles ou égales<sup>(1)</sup> :

$$f(x) = \frac{1}{\sqrt{2\pi}} \left[ \frac{p_1}{\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{p_2}{\sigma_2} e^{-\frac{x^2}{2\sigma_2^2}} \right]$$

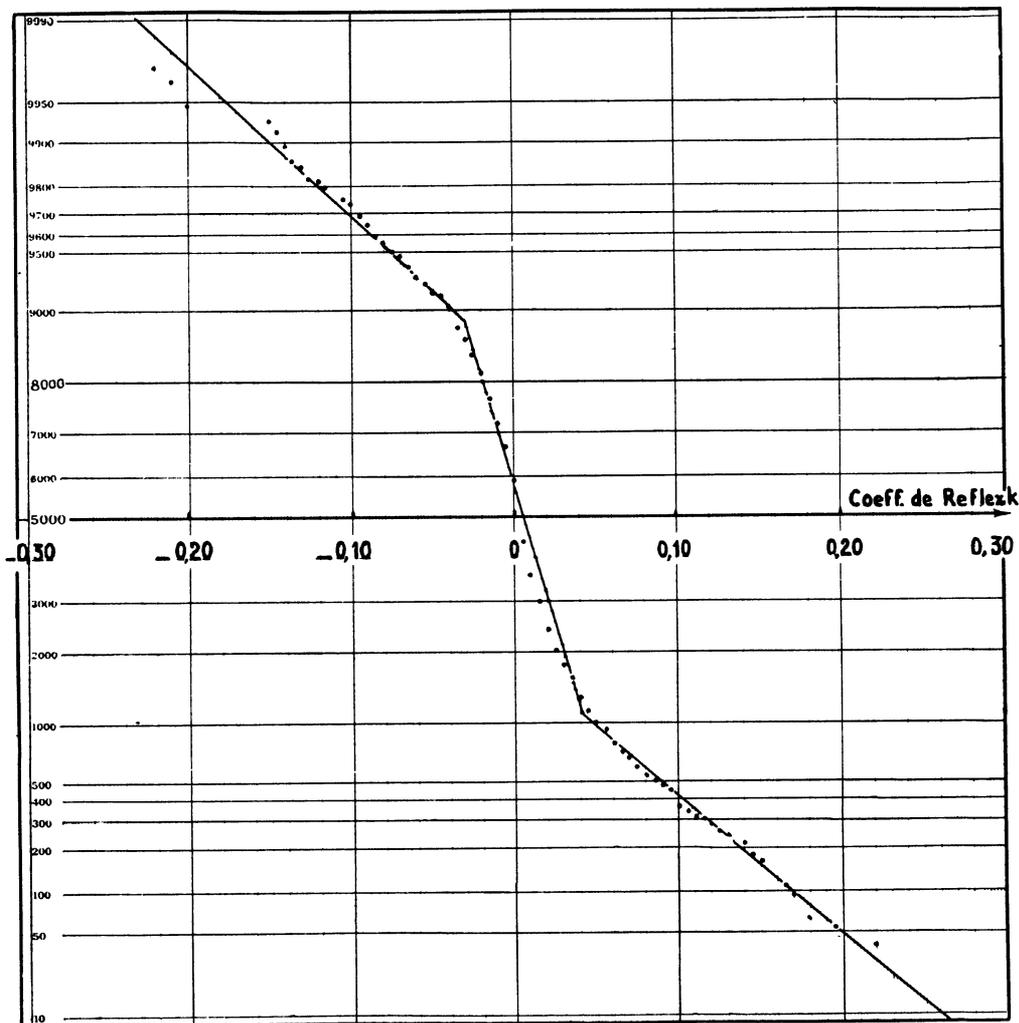


Figure 1 - Ajustement à une loi normale des coefficients de réflexion.  
734 observations. dt = 1 ms.

(1) Le calcul de  $E(X)$  permet de se ramener au cas où  $m = 0$  car  $E(X) = mp_1 + mp_2 = m$ .

On peut établir facilement les relations :

$$p_1 + p_2 = 1 \quad (1)$$

$$E(|X|) = \sqrt{\frac{2}{\pi}} (p_1 \sigma_1 + p_2 \sigma_2) = A \quad (2)$$

$$E(X^2) = p_1 \sigma_1^2 + p_2 \sigma_2^2 = B \quad (3)$$

$$E(|X|^3) = \frac{4}{\sqrt{2\pi}} (p_1 \sigma_1^3 + p_2 \sigma_2^3) = C \quad (4)$$

où A, B, C sont les moments empiriques obtenus à partir des observations.

Les quatre équations précédentes doivent nous permettre d'estimer  $p_1$ ,  $p_2$ ,  $\sigma_1$ ,  $\sigma_2$ , à partir de  $a = A\sqrt{\frac{\pi}{2}}$ ,  $b = B$  et  $c = \sqrt{\frac{2\pi}{16}} C$ .

$$p_1 + p_2 = 1 \quad (5)$$

$$p_1 \sigma_1 + p_2 \sigma_2 = a \quad (6)$$

$$p_1 \sigma_1^2 + p_2 \sigma_2^2 = b \quad (7)$$

$$p_1 \sigma_1^3 + p_2 \sigma_2^3 = c \quad (8)$$

Le produit de la relation (5) par  $\sigma_1$  moins la relation (6) donne  $p_2(\sigma_1 - \sigma_2) = \sigma_1 - a$ .

En procédant de même avec les relations (6) et (7), et (7) et (8) on obtient trois relations :

$$p_2(\sigma_1 - \sigma_2) = \sigma_1 - a \quad (9)$$

$$p_2 \sigma_2(\sigma_1 - \sigma_2) = a\sigma_1 - b \quad (10)$$

$$p_2 \sigma_2^2(\sigma_1 - \sigma_2) = b\sigma_1 - c \quad (11)$$

Le rapport de (11) sur (10) et de (10) sur (9) donne :

$$\sigma_2 = \frac{b\sigma_1 - c}{a\sigma_1 - b} \quad (12)$$

$$\sigma_2 = \frac{a\sigma_1 - b}{\sigma_1 - a} \quad (13)$$

L'égalisation de (12) et (13) donne l'expression de  $\sigma_1$  :

$$(a\sigma_1 - b)^2 = (b\sigma_1 - c)(\sigma_1 - a) \\ \sigma_1^2(a^2 - b) - (ab - c)\sigma_1 + b^2 - ac = 0 \quad (14)$$

On obtiendrait la même équation en  $\sigma_2$  en multipliant (5), (6) et (7) par  $\sigma_2$  au lieu de  $\sigma_1$ .

On a donc :

$$\sigma = \frac{ab - c \pm \sqrt{(ab - c)^2 - 4(a^2 - b)(b^2 - ac)}}{2(a^2 - b)}$$

Les racines de (14) donnent les valeurs de  $\sigma_1$  et  $\sigma_2$ .

Ces racines existent et sont réelles car :

$$(ab - c)^2 - 4(a^2 - b)(b^2 - ac) > 0$$

En effet :

$$b^2 - ac = p_1 p_2 \sigma_1 \sigma_2 (2 \sigma_1 \sigma_2 - \sigma_1^2 - \sigma_2^2) = p_1 p_2 \sigma_1 \sigma_2 (\sigma_1 - \sigma_2)^2 < 0$$

$$a^2 - b = (p_1^2 - p_1) \sigma_1^2 + (p_2^2 - p_2) \sigma_2^2 + 2 p_1 p_2 \sigma_1 \sigma_2 = p_1 p_2 (\sigma_1^2 + \sigma_2^2 + 2 \sigma_1 \sigma_2) > 0$$

Donc :

$$- 4(a^2 - b)(b^2 - ac) > 0$$

Les expressions de  $p_1$  et  $p_2$  s'obtiennent aisément en reportant les valeurs de  $\sigma_1$  et  $\sigma_2$  dans (9), puis (5). Les résultats fournis par ces équations nous semblent cependant exiger un grand nombre d'observations, car la dispersion existant sur  $a$ ,  $b$  et  $c$  entraîne une dispersion très grande sur les estimations de  $\sigma_1$ ,  $\sigma_2$ ,  $p_1$  et  $p_2$ .

C'est pourquoi l'on peut se demander si l'on n'a pas intérêt à estimer ces paramètres à partir du graphique de Henri en combinant dans diverses proportions deux lois normales d'écart-types différents et en se rapprochant par approximations successives de la loi empirique observée.

Nous donnons (figure 2) un exemple de combinaison de deux lois normales sur le graphique de Henri. Comme la droite de Henri théorique de chaque loi indique pour une abscisse  $t$  :

$$G(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx$$

on voit que, connaissant  $G_1(t)$  et  $G_2(t)$  ainsi que  $p$ , on peut calculer facilement la probabilité  $G(t)$  pour que la variable aléatoire issue du mélange soit supérieure à  $t$  :

$$G(t) = pG_1(t) + (1 - p) G_2(t)$$

Le graphique théorique de la figure 2 montre trois alignements joints par des lignes un peu incurvées. Il convient du reste de remarquer que les trois tronçons de la courbe  $G(t)$  ne sont pas tout à fait rectilignes sur le graphique de Henri, mais il s'en faut de peu.

Pour étudier les résultats de cette méthode, nous avons procédé à des essais sur des "populations synthétiques". Pour cela, nous mélangeons dans une certaine proportion deux populations normales issues de tirages aléatoires avec même moyenne et des  $\sigma$  donnés.

Les paramètres théoriques  $p_1$ ,  $p_2$ ,  $\sigma_1$  et  $\sigma_2$  sont ainsi connus a priori. Les valeurs exactes de  $\hat{\sigma}_1$ ,  $\hat{\sigma}_2$ ,  $\hat{m}_1$  et  $\hat{m}_2$  dans les échantillons des deux sous-populations sont également connus ( $\hat{m}_1$  et  $\hat{m}_2$  peuvent ne pas être identiques et les dispersions des échantillons peuvent différer des dispersions théoriques  $\sigma_1$  et  $\sigma_2$ ).

Ayant constitué la population composée, nous lui appliquons la méthode d'estimation précédente. Sur le tableau suivant, nous résumons les résultats de nos calculs sur 3 exemples, dont nous fournissons la représentation graphique en papier gaussien sur les figures 3, 4, 5.

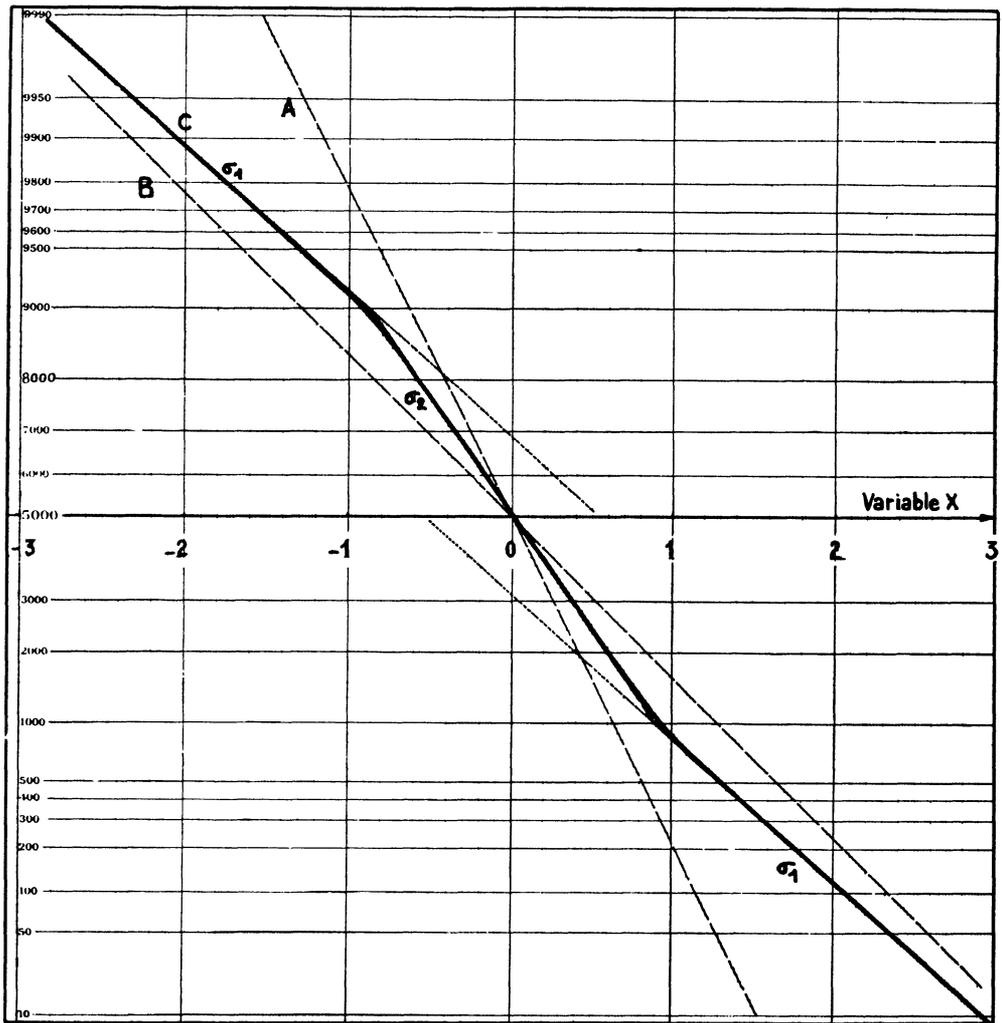
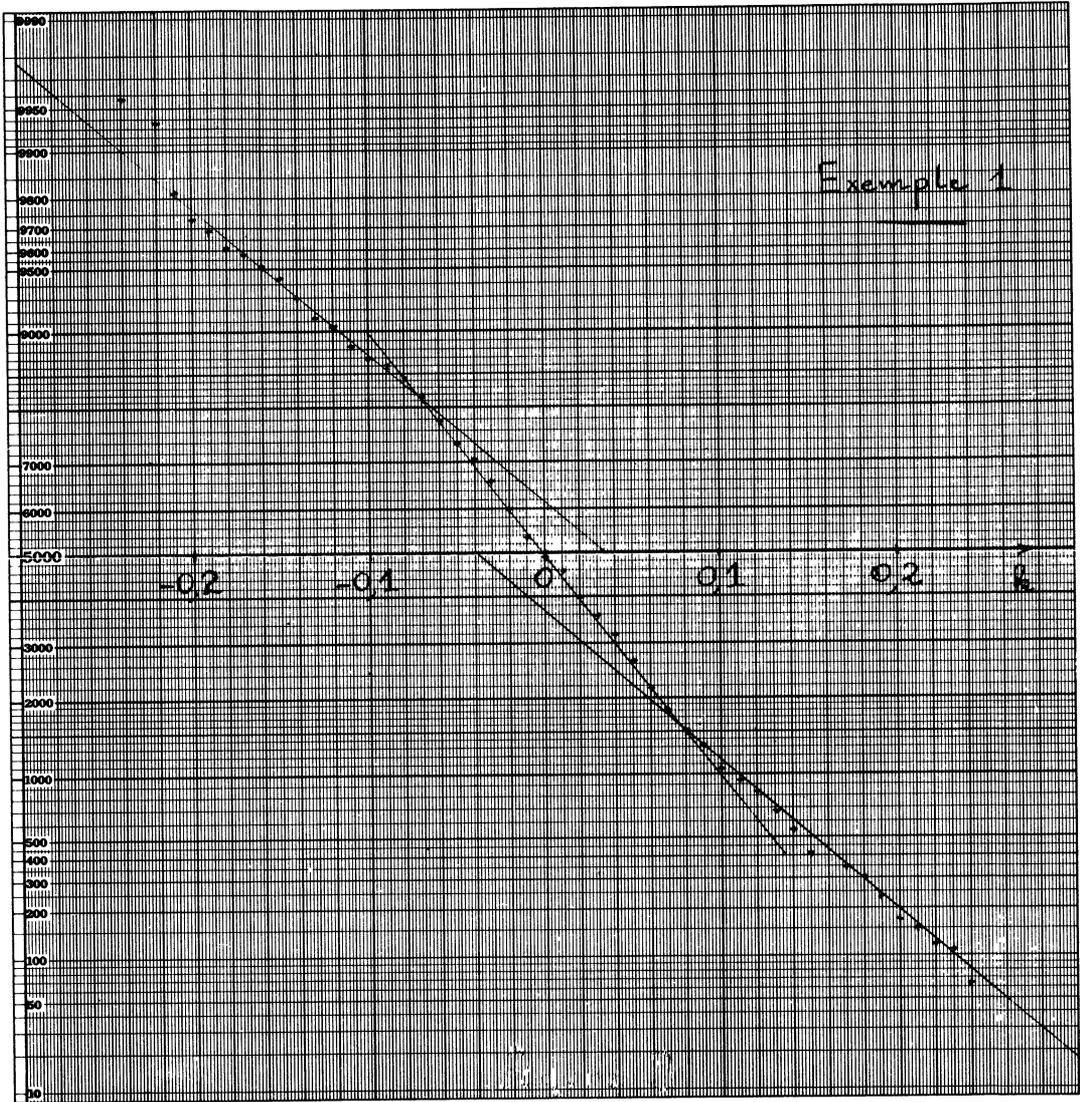


Figure 2 - Composition à égalité de deux lois normales de moyennes nulles et d'écart-types égaux à 1 et  $\frac{1}{2}$ . A - Population  $m = 0$ ,  $\sigma = \frac{1}{2}$ . B - Population  $m = 0$ ,  $\sigma = 1$ .

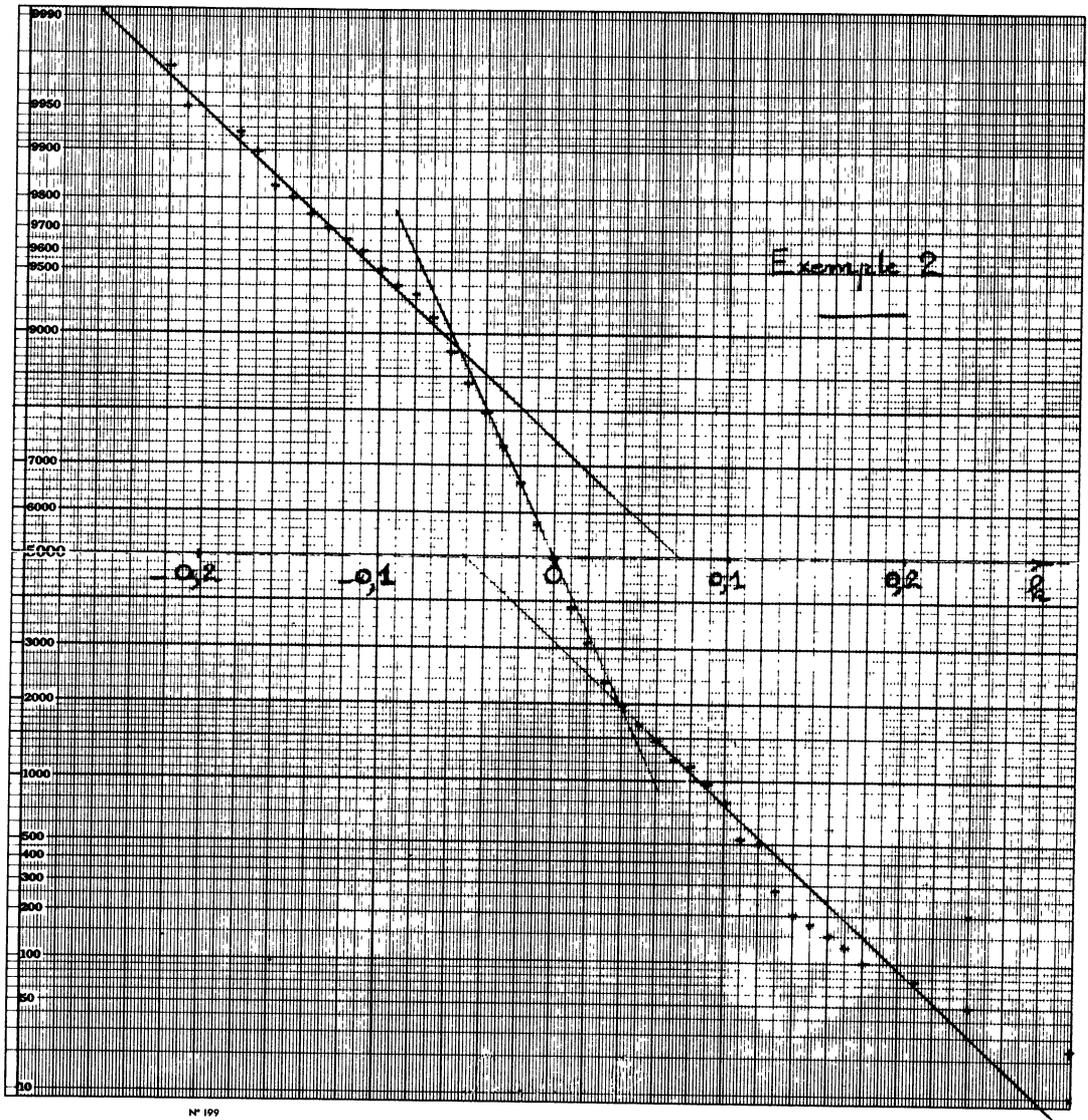
C - Population résultante :  $m = 0$ ,  $\begin{cases} \sigma_1 = 1,1 \\ \sigma_2 = 0,7 \end{cases}$ .

L'estimation par le calcul des moments absolus d'ordre 1, 2 et 3 nécessite des calculs précis avec de nombreuses décimales. En particulier il convient de vérifier soigneusement la nullité de la moyenne et d'effectuer les calculs sur les moments de  $|x - \bar{x}|$  si  $\bar{x}$  n'est pas strictement nul. Faute de quoi les estimations  $\hat{\sigma}_1$  et  $\hat{\sigma}_2$  risquent d'être complètement fausses, voire imaginaires. D'autre part les classes  $x_{i+1} - x_i$  des observations doivent être assez fines, sinon des corrections du genre de celle de Sheppard sont nécessaires. En dépit de ces précautions, les résultats obtenus dans les 3 exemples précédents ne sont pas très satisfaisants.



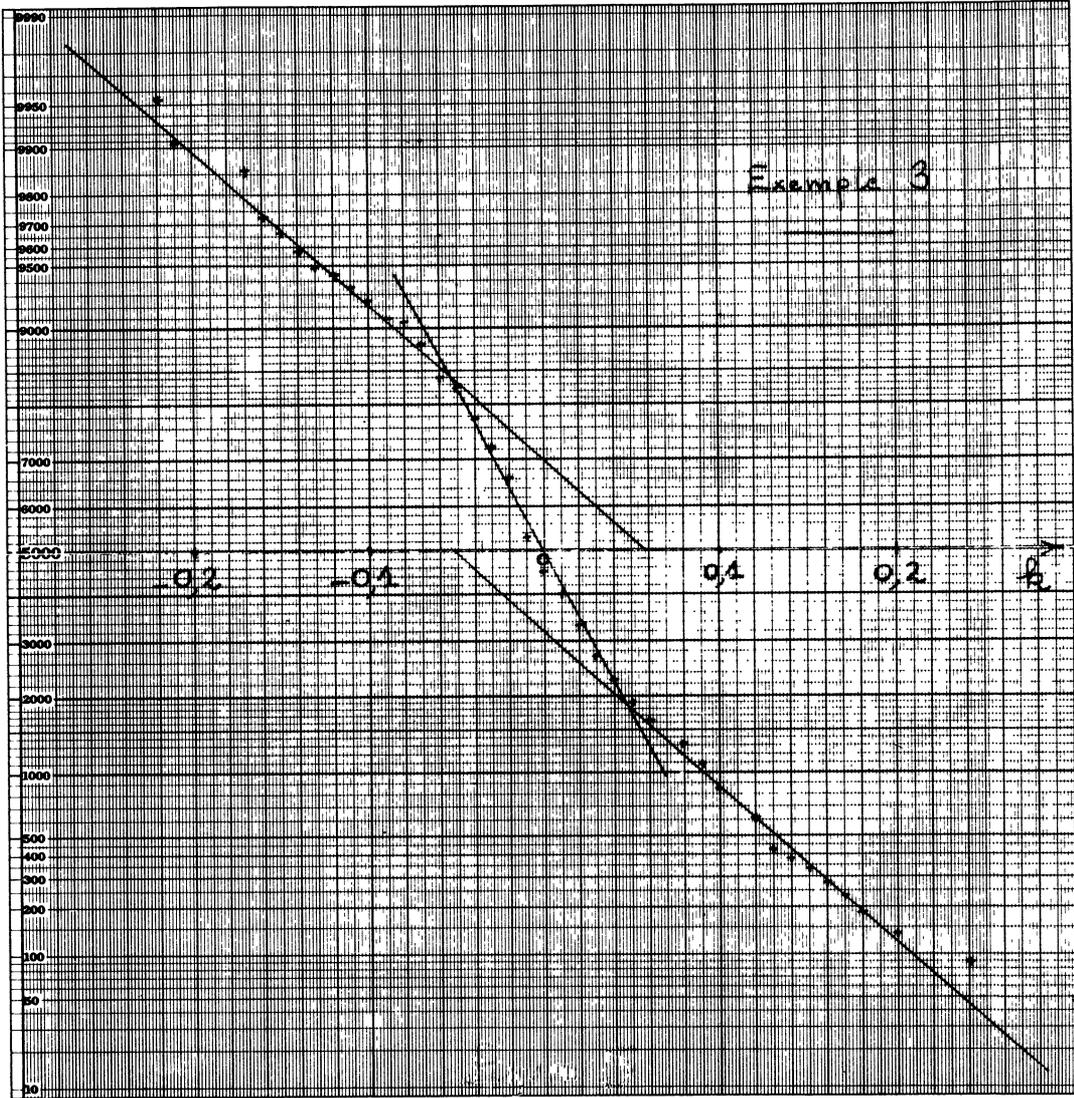
N° 199

Figure 3



N° 199

Figure 4



N° 199

Figure 5

Expérience		1		2		3	
Nbre d'observ.		454		400		210	
Données théoriques	$\sigma_1$	0.05		0.04		0.04	
	$\sigma_2$	0.10		0.10		0.10	
	$p_1$	34 %		75 %		52 %	
	$p_2$	66 %		25 %		48 %	
Dispersion de l'échantillon	$\hat{\sigma}_1$	0.051		0.042		0.035	
	$\hat{\sigma}_2$	0.103		0.102		0.102	
Estimations par le calcul	$\sigma_1$	0.036	(1) -28%	0.032	(1) -20%	0.041	(1) + 3%
	$\sigma_2$	0.096	- 4%	0.092	- 8%	0.105	+ 5%
	$p_1$	17%	-48%	57%	-24%	60%	+15%
	$p_2$	83%	+24%	43%	+72%	40%	-17%
Estimations directes par le graphique	$\sigma_1$	0.080	+60%	0.044	+ 8%	0.053	+32%
	$\sigma_2$	0.114	+14%	0.104	+ 4%	0.112	+12%
	$p_1$	18%	-47%	76%	+ 2%	76%	+46%
	$p_2$	82%	+24%	24%	- 4%	24%	-50%
(1) Erreur relative en pourcentage par rapport aux données théoriques.							

On peut également employer une méthode graphique à partir du graphique de Henri. Les pentes des deux alignements extrêmes donnent une estimation par excès de l'écart-type  $\sigma_2$ . La pente de l'alignement central donne une estimation par excès de  $\sigma_1$ . L'estimation de  $\hat{\sigma}$  est obtenue facilement en prenant  $\hat{\sigma} = \frac{k(84\%) - k(16\%)}{2}$  où  $k(x\%)$  est l'abscisse telle que  $x\%$  des observations lui soient supérieures, lue directement sur la droite de Henri (voir (2)). Le pourcentage de cette dernière population peut être obtenu à l'aide du calcul de  $E|x - \bar{x}| \sqrt{\frac{\pi}{2}} = a$  et des estimations précédentes  $\hat{\sigma}_1$  et  $\hat{\sigma}_2$  en appliquant la formule :

$$p_1 = \frac{\hat{\sigma}_2 - a}{\hat{\sigma}_2 - \hat{\sigma}_1}$$

(obtenue à partir de la combinaison des relations (5) et (6)).

On peut remarquer que ces calculs surestiment les valeurs des  $\sigma$ , mais que les erreurs commises ne sont guère plus élevées que par la méthode des moments en valeur absolue.

Une étude de la précision à attendre de ces estimations, en fonction des valeurs des paramètres  $\sigma_1$ ,  $\sigma_2$ ,  $p_1$  et  $p_2$  serait du reste à faire.

Signalons que la discrimination de deux populations normales mélangées a été étudiée par quelques auteurs dans le cas où les moyennes sont différentes (3), (5), (6), (7). On a alors 6 paramètres. Pearson propose un calcul à partir des 5 premiers moments de la distribution empirique (voir (5) ou (6)). Les solutions sont parfois indéterminées et la dispersion sur les estimations est certainement très grande. Lorsque les moyennes  $m_1$  et  $m_2$  sont connues, on peut en déduire facilement les proportions  $p_1$  et  $p_2$  car  $E(X) = p_1 m_1 + m_2(1 - p_1)$ . On peut aussi procéder par approximations successives à partir de la solution de Pearson (voir (6)).

### III - ETUDE DE QUELQUES FONCTIONS DE VARIABLES NORMALES -

Nous avons vu au paragraphe 2 que les coefficients de réflexion semblaient composés du mélange de deux populations normales. Si nous admettons que les vitesses de propagation du son  $V(h)$  aux diverses profondeurs suivent une loi normale avec une moyenne et un écart-type pratiquement stationnaires, mais une légère corrélation entre  $V(h)$  et  $V(h + dh)$  pratiquement indépendante de  $h$ , on peut se demander si la variable

$$k(h) = \frac{V(h) - V(h + dh)}{V(h) + V(h + dh)}$$

n'aurait pas une distribution analogue à celles observées au paragraphe 2. C'est pourquoi nous analysons ici la répartition de quelques fonctions de deux variables normales corrélées  $x$  et  $y$  (voir (1)).

Les fonctions analysées ici sont :

- la somme  $x + y$ ;
- la différence  $x - y$ ;
- le couple  $x + y, x - y$ ;
- le produit  $xy$ ;
- le rapport  $\frac{x}{y}$  ;
- le rapport  $\frac{x - y}{x + y}$ .

a) Ces quelques fonctions se rencontrent en particulier dans le cas d'une série temporelle  $x(t), y(t) = x(t + h)$  où  $x(t)$  peut être considéré comme une variable normale à autocorrélation non nulle. Par exemple, si  $x(t)$  est la vitesse de propagation d'un milieu, on s'intéresse à la loi de répartition des coefficients de réflexion :  $\frac{x(t) - x(t + h)}{x(t) + x(t + h)}$ , sachant que  $x(t)$  est une loi normale dont l'autocorrélation est connue. Ces problèmes se rencontrent aussi dans le cas du radar monopulse (4).

b) Loi de  $x + y$ . La densité de probabilité du couple  $(x, y)$  est :

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{(x-m)^2 + (y-m)^2 - 2r(x-m)(y-m)}{2\sigma^2(1-r^2)}}$$

On suppose dans tout ce qui suit que  $x$  et  $y$  ont même moyenne  $m$ , même variance  $\sigma^2$  et une corrélation  $r$ .

Pour déterminer la densité de probabilité d'une fonction  $u = g(x, y)$ , on fait en général un changement de variable pour passer du couple  $(x, y)$  au couple  $(u, x)$  ou  $(u, y)$  ou  $(u, v)$  avec  $v = h(x, y)$ . On intègre ensuite par rapport à la seconde variable et on obtient la densité de probabilité de la variable  $u$ .

Dans le cas de la somme  $u = x + y$ , on retrouve le résultat bien connu :  $u$  est une variable normale.

Le couple (u, x) a pour densité de probabilité :

$$\frac{1}{2 \pi \sigma^2 \sqrt{1 - r^2}} e^{-\frac{(x-m)^2 + (u-x-m)^2 - 2r(x-m)(u-x-m)}{2\sigma^2(1-r)^2}} dx du$$

$$= \frac{1}{2 \pi \sigma^2 \sqrt{1 - r^2}} e^{-\frac{2(1+r)(x-\frac{u}{2})^2 + (1-r)(\frac{u}{2} + 2m^2 - 2mu)}{2\sigma^2(1-r^2)}} dx du$$

En intégrant cette expression par rapport à x, ce qui est possible par un nouveau changement de variable ( $t = x - \frac{u}{2}$ ), on obtient pour loi de u :

$$f(u) = \frac{1}{2 \sigma \sqrt{(1+r)} \pi} e^{-\frac{(u-2m)^2}{4\sigma^2(1+r)}}$$

La variable  $u = x + y$  est une variable normale de moyenne  $2m$  et d'écart-type  $\sigma \sqrt{2(1+r)}$ . Quand  $r = 0$ , on retrouve le résultat classique d'addition de deux variables indépendantes.

c) Loi de x - y. Par un calcul analogue au précédent, on obtient pour densité de probabilité de la différence  $u = x - y$

$$f(u) = \frac{1}{2 \sigma \sqrt{(1-r)} \pi} e^{-\frac{u^2}{4\sigma^2(1-r)}}$$

La différence u est une variable normale de moyenne nulle et d'écart-type  $\sigma \sqrt{2(1-r)}$ .

d) Loi du couple x - y, x + y. Effectuons le changement de variable  $\begin{cases} u = x + y \\ v = x - y \end{cases}$ .

La densité de probabilité du couple (u, v) est, avec le Jacobien  $\frac{D(u, v)}{D(x, y)} = \frac{1}{2}$  et  $x = \frac{u+v}{2}$ ,  $y = \frac{u-v}{2}$  :

$$\frac{1}{4 \pi \sigma^2 \sqrt{1 - r^2}} e^{-\frac{(\frac{u+v}{2}-m)^2 + (\frac{u-v}{2}-m)^2 - 2r(\frac{u+v}{2}-m)(\frac{u-v}{2}-m)}{2\sigma^2(1-r^2)}} = \frac{1}{4 \pi^2 \sigma^2 \sqrt{1 - r^2}} e^{-\frac{(u-2m)^2(1-r) + v^2(1+r)}{4\sigma^2(1-r^2)}}$$

On retrouve les résultats exposés aux paragraphes b et c : loi de  $x + y$  et loi de  $x - y$ . De plus, on remarque que u et v sont deux variables aléatoires indépendantes, car leurs densités se séparent dans la densité du couple (u, v). Si les valeurs moyennes ou les écarts-types des deux variables x et y ne se trouvaient pas égaux, les termes rectangles en uv ne disparaîtraient pas et u et v seraient deux variables normales corrélées.

e) Loi du produit xy. La densité de probabilité du couple  $xy = u$  et de  $x$  est, dans le cas de moyenne nulle de  $x$  et  $y$  :

$$\frac{1}{2\pi\sigma^2\sqrt{1-r^2}|x|} e^{-\left(x^2 + \frac{u^2}{x^2} - 2ru\right) \frac{1}{2\sigma^2(1-r^2)}} \left( \frac{D(u, v)}{D(x, y)} = \left| \frac{1}{x} \right| \right)$$

La densité de probabilité de  $u$  s'obtient par intégration sur  $x$  :

$$\int_{-\infty}^{\infty} e^{-\left(x^2 + \frac{u^2}{x^2} - 2ru\right) \frac{1}{2\sigma^2(1-r^2)}} \frac{1}{2\pi\sigma^2\sqrt{1-r^2}|x|} dx = \frac{e^{-\frac{ru}{\sigma^2(1-r^2)}}}{\pi\sigma^2\sqrt{1-r^2}} k_0\left(\frac{|u|}{1-r^2}\right)$$

où  $k_0$  est la fonction de Bessel Hankel d'ordre 0.

Si les variables  $x$  et  $y$  sont indépendantes ( $r = 0$ ), on a :

$$f(u) = \frac{1}{\pi\sigma^2} k_0(|u|)$$

Si la corrélation  $r = 1$ , la loi de  $u$  est une loi du type III de Pearson :

$$f(u) = \frac{e^{-\frac{|u|}{2}}}{\sqrt{2}\pi\sigma|u|}$$

En utilisant les tables de Bessel, on peut calculer  $f(u)$  pour diverses valeurs de  $r$ . Morlat a construit quelques courbes de ce type pour  $r = 0; 0,5; 0,9; 1$ .

f) Loi du rapport  $\frac{x}{y}$ . La densité de probabilité du couple  $\frac{x}{y} = u$  et de  $y$  est obtenue à partir de la densité du couple  $(x, y)$  du paragraphe b) avec un Jacobien  $\frac{D(u, y)}{D(x, y)} = |y|$ .

$$\frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\left[(uy-m)^2 + (y-m)^2 - 2r(uy-m)(y-m)\right] \frac{1}{2\sigma^2(1-r^2)}} |y| dy du =$$

$$= \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\left[\left(y\sqrt{u^2-2ru+1} - \frac{m(1-r)(1+u)}{\sqrt{u^2-2ur+1}}\right)^2 + \frac{(1-r^2)(u-1)^2 m^2}{u^2-2ru+1}\right] \frac{1}{2\sigma^2(1-r^2)}} |y| dy du$$

$$f(u) = \frac{e^{-\frac{(u-1)^2 m^2}{(u^2-2ru+1)2\sigma^2}}}{2\pi\sigma^2\sqrt{1-r^2}} \int_{-\infty}^{\infty} e^{-\left(y\sqrt{u^2-2ru+1} - \frac{m(1-r)(u+1)}{\sqrt{u^2-2ur+1}}\right)^2 \frac{1}{2\sigma^2(1-r^2)}} |y| dy$$

Le calcul de  $\int_{-\infty}^{\infty} e^{-(y-a)^2 b} |y| dy$  donne :

$$\int_0^{\infty} e^{-(y-a)^2 b} y dy + \int_0^{\infty} e^{-(y+a)^2 b} y dy = \int_0^{\infty} e^{-(y-a)^2 b} (y-a+a) dy + \int_0^{\infty} e^{-(y+a)^2 b} (y+a-a) dy$$

$$\int_a^\infty e^{-xb} \frac{dX}{2} + a \int_{-a}^\infty e^{-y^2b} dY + \int_a^\infty e^{-x'b} \frac{dX'}{2} - a \int_a^\infty e^{-y'^2b} dY'$$

avec  $X = (y - a)^2$ ;  $Y = y - a$ ;  $X' = (y + a)^2$ ;  $Y' = y + a$

$$= \frac{e^{-ba^2}}{2b} + \frac{a\sqrt{2\pi}}{\sqrt{2b}} G(-a) + \frac{e^{-ba^2}}{2b} - \frac{a\sqrt{2\pi}}{\sqrt{2b}} G(a) = \frac{e^{-ba^2}}{b} + a\sqrt{\frac{\pi}{b}} [G(-a) - G(a)]$$

(où  $G(a) = \frac{\sqrt{2b}}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{x^2 2b}{2}} dx$  est la loi de Gauss de variance  $\frac{1}{b}$  et

$$A = e^{-\frac{m^2(1-r)(u+1)^2}{2\sigma^2(1+r)(u^2-2ur+1)}}).$$

$$f(u) = \frac{e^{-\frac{(u-1)^2 m^2}{2\sigma^2(u^2-2ur+1)}}}{2\pi\sigma^2\sqrt{1-r^2}} \left[ \frac{2\sigma^2(1-r^2)A}{u^2-2ur+1} + \frac{m(1-r)(u+1)}{u^2-2ur+1} \sqrt{\frac{2\pi\sigma^2(1-r^2)}{u^2-2ur+1}} \Pi \right]$$

avec

$$\Pi = \int_{-\frac{m(1-r)(u+1)}{u^2-2ur+1}}^{\frac{m(1-r)(u+1)}{u^2-2ur+1}} e^{-\frac{x^2}{2} \frac{[u^2-2ur+1]}{\sigma^2(1-r^2)}} \frac{dx}{\sqrt{2\pi} \sqrt{\frac{\sigma^2(1-r^2)}{u^2-2ur+1}}}$$

Lorsque  $m = 0$ , on trouve la loi de Cauchy :

$$f(u) = \frac{\sqrt{1-r^2}}{(u^2-2ur+1)\pi}$$

la variable réduite est  $\frac{u-r}{\sqrt{1-r^2}}$ .

Si  $r = 1$ , on a toujours  $u = 1$ .

Si  $m$  devient très grand par rapport à  $\sigma$ ,  $\Pi$  et  $A$  sont pratiquement égaux à 1 et  $u$  est distribué pratiquement comme une loi normale de moyenne 1 et d'écart-type  $\frac{\sigma\sqrt{2}}{m\sqrt{1-r}}$ .

g) Loi du rapport  $\frac{x-y}{x+y}$ . On recherche la densité de probabilité du couple  $(u = \frac{x-y}{y+y}, y)$  à partir de celle du couple  $(x, y)$ . On a :

$$\frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-[(y\frac{1+u}{1-u}-m)^2 + (y-m)^2 - 2r(y\frac{1+u}{1-u}-m)(y-m)]} \frac{1}{2\sigma^2(1-r^2)} \frac{2|y|}{(1-u)^2} dy du$$

(le Jacobien est égal à  $\frac{2|y|}{(1-u)^2}$ ).

On peut transformer cette expression en :

$$\frac{1}{2\pi\sigma^2\sqrt{1-r^2}} \frac{e^{-\frac{m^2 u^2}{\sigma^2(1+u^2-r(1-u^2))}}}{(1-u^2)} \int_{-\infty}^{\infty} e^{-\left(y - \frac{m(1-r)(1-u)}{u^2+1-r(1-u^2)}\right)^2 \frac{u^2+1-r(1-u^2)}{\sigma^2(1-r^2)(1-u^2)}} 2|y| dy du$$

Il reste à effectuer un calcul analogue à celui du paragraphe f) : loi du rapport  $\frac{x}{y}$ , et on trouve pour densité  $f(u)$  :

$$f(u) = \frac{1}{\pi \sigma^2 \sqrt{1-r^2}} \frac{e^{-\frac{m^2 u^2}{1+u^2-r(1-u^2)}}}{(1-u)^2} \left[ \frac{\sigma^2(1-r^2)(1-u)^2}{u^2+1-r(1-u^2)} A' \right. \\ \left. + \frac{\prod m(1-r)(1-u)}{u^2+1-r(1-u^2)} \sqrt{\frac{\pi \sigma^2(1-r)(1-u^2)}{u^2+1-r(1-u^2)}} \right]$$

$$\prod = \int_{-\frac{m(1-r)(1-u)}{u^2+1-r(1-u^2)}}^{\frac{m(1-r)(1-u)}{u^2+1-r(1-u^2)}} \frac{e^{-\frac{x^2}{\sigma^2(1-r^2)(1-u)^2}}}{\sqrt{2\pi} \sqrt{\frac{\sigma^2(1-r^2)(1-u)^2}{2(u^2+1-r(1-u^2))}}} dx, \quad A' = e^{-\frac{m^2(1-r)}{\sigma^2(1+r)[u^2+1-r(1-u^2)]}}$$

Si  $m = 0$ , on retrouve la loi de Cauchy :

$$f(u) = \frac{1+r}{\pi \left( u^2 \frac{1+r}{1-r} + 1 \right)}$$

La variable réduite est  $u \sqrt{\frac{1+r}{1-r}}$ .

Si  $r = 1$ , la variable  $u = 0$ .

Si  $m$  est très grand devant  $\sigma$ ,  $\prod$  et  $A'$  sont pratiquement égaux à 1 et  $u$  est pratiquement distribué comme une loi normale de moyenne 0 et d'écart-type  $\frac{\sigma \sqrt{1-r}}{m \sqrt{2}}$ .

Vérification sur un exemple synthétique - Pour vérifier les considérations précédentes sur un exemple simple, nous avons constitué un échantillon de vitesses normales  $V(t)$  de moyenne 3 500 m/s et d'écart-type  $\sigma_v = 950$  m/s. Nous avons tiré 300 vitesses au hasard suivant cette loi (avec une corrélation nulle entre les  $V(t)$ ). La figure 6 montre la droite de Henri théorique et les points représentatifs de l'échantillonnage. A partir de cet échantillonnage, on a construit 299 valeurs  $k$  par la formule :

$$k_j = \frac{V_j - V_{j+1}}{V_j + V_{j+1}}$$

On constate que les  $k_j$  peuvent être assimilés à une loi normale sur le graphique de la figure 7.

La moyenne des  $k(t)$  dans les deux cas est nulle et leur écart-type est de 0,20 et 0,22 alors que l'on devrait avoir  $\sigma = \frac{\sqrt{2} \cdot \sigma_v}{2m} = \frac{\sqrt{2} \cdot 950}{2 \cdot 3500} = 0,192$  d'après les formules théoriques précédentes. L'accord est très satisfaisant.

Enfin, en translatant toutes les vitesses de l'échantillon précédent pour leur donner une vitesse moyenne nulle et en calculant les  $k_j = \frac{V_j - V_{j+1}}{V_j + V_{j+1}}$  on obtient un ajustement excellent à une loi de Cauchy. Les formules théoriques sont donc bien vérifiées.

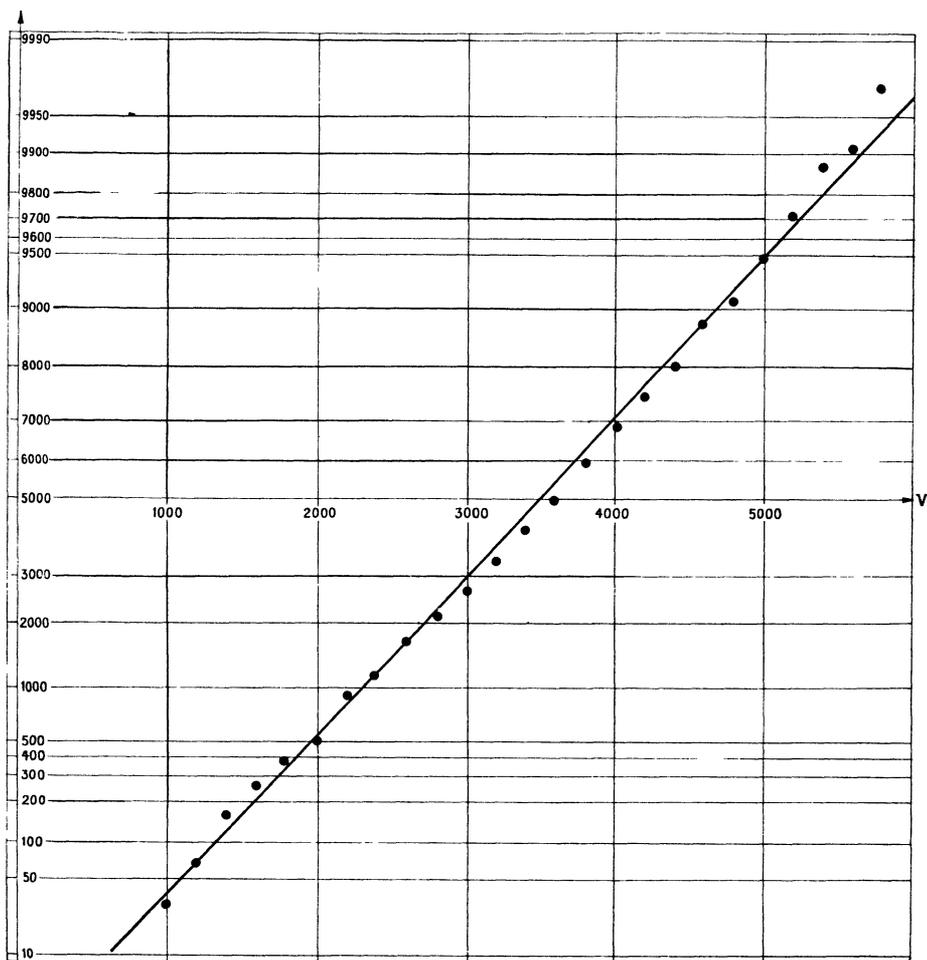


Figure 6 - Echantillon de vitesses synthétiques suivant la loi normale.  $m = 3\ 500$ .  
 $\sigma = 900$ .  $n = 300$ .

Avec les dispersions relevées sur les  $V(h)$ , la loi résultante de  $k(h)$  doit tendre vers une loi normale et l'existence des 3 alignements constatée sur le graphique de Henri ne résulte donc pas de la formule  $\frac{V(h) - V(h + dh)}{V(h) + V(h + dh)}$  mais elle est sans doute provoquée par la composition de couches géologiques hétérogènes ayant des dispersions variables.

#### BIBLIOGRAPHIE

- [1] AGARD - Thèse de Doctorat sur l'Analyse Statistique et probabiliste des Sismogrammes. (Revue de l'Institut Français du Pétrole. Vol. XVI. n° 3 et 4. 1961).

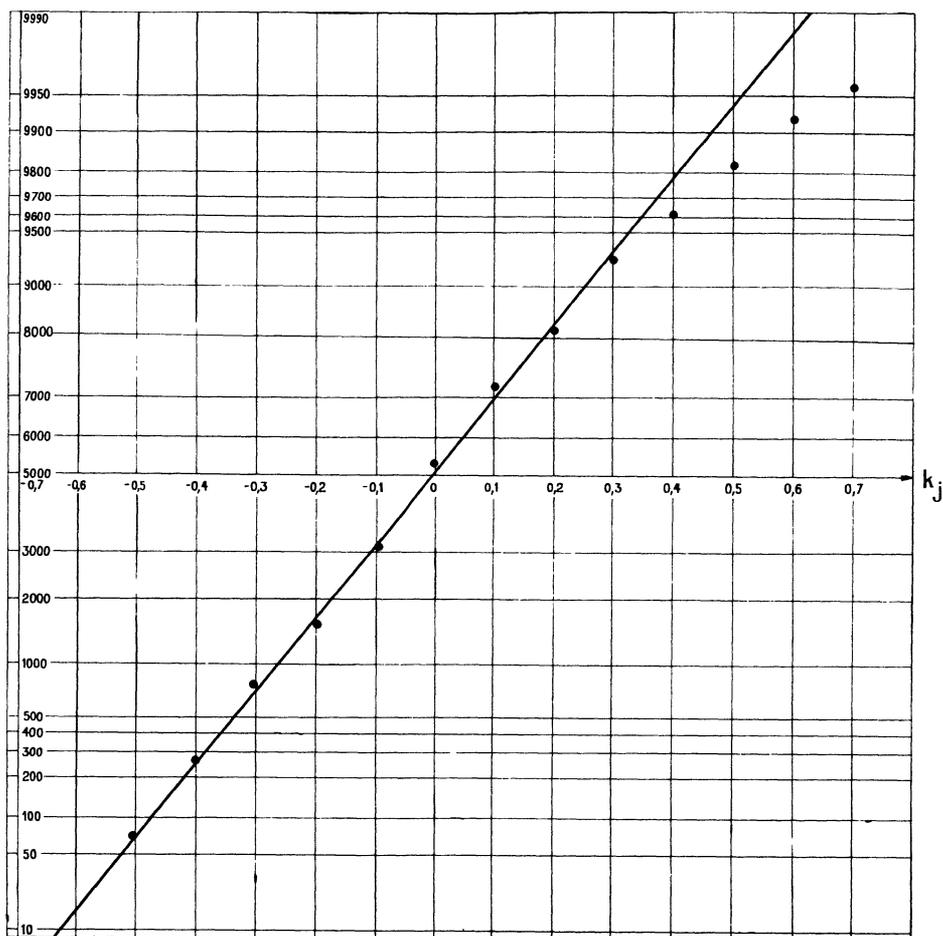


Figure 7 - Ajustement à la loi normale des coefficients de réflexion synthétiques.

$$k_j = \frac{V_j - V_{j+1}}{V_j + V_{j+1}} \quad m = 0. \quad \sigma = 0,21. \quad n = 299.$$

- [2] AGARD - Thèse de Statistique sur l'Application de la loi log-normale à la Recherche Minière. Institut de Statistique de l'Université de Paris. 1956.
- [3] BURRAU - The half invariants of two typical laws of errors with an application to the problem of dissecting a frequency curve into components. (Skand. Aktuarietiedschrift 17. 1934).
- [4] FORTET - Sur la détermination du spectre de l'inverse d'une fonction aléatoire et ses applications. (Publications de l'I. S. U. P. vol. VI, fasc. 3. 1957).
- [5] PEARSON - Philosophical Transactions vol.185. part.I. page 13.
- [6] RISSER TRAYNARD - Les principes de la statistique mathématique. Gauthier-Villars. 1933. p. 96.

- [7] STROMGEN - Tables and diagrams for dissecting a frequency curve into components. (Skand Aktuarietiedschrift, 1934. 17. 7. 54).
- [8] J. AGARD et G. GRAU - Etude statistique de sismogrammes geophysical Prospecting (en cours de publication).