

Y. BRUNET-MORET \*

# Étude de quelques lois statistiques utilisées en hydrologie

Nous n'avons pas l'intention d'écrire un traité de probabilité ou statistique mathématique, ni même des études exhaustives de lois de probabilités, mais plus simplement de proposer aux utilisateurs des moyens, qui pourront être perfectionnés, de se servir plus facilement de lois déjà introduites en hydrologie statistique par de nombreux auteurs.

Les méthodes envisagées le sont surtout sous l'angle du calcul par ordinateur, pour :

- la loi de GAUSS ou loi normale;
- la loi de GUMBEL ou loi doublement exponentielle;
- la loi de GALTON ou loi log normale ou gausso-logarithmique;
- la loi de PEARSON III ou loi gamma incomplète;
- la loi exponentielle généralisée (ou de FRÉCHET, de GOODRICH, de JENKINSON);
- la loi de PEARSON I ou loi bêta incomplète.

---

\* Ingénieur Hydrologue de l'O.R.S.T.O.M.

# SOMMAIRE

I. — Généralités. . . . .	5
1.1. Définitions. . . . .	5
1.2. Paramètres . . . . .	5
1.3. Valeurs centrales — Moments . . . . .	6
1.4. Lois tronquées et troncatures . . . . .	9
1.5. Moments déduits d'un échantillon de taille connue . . . . .	10
1.6. Mélange de distributions. . . . .	12
II. — Détermination des paramètres . . . . .	13
2.1. . . . .	13
2.2. Qualificatifs des estimations . . . . .	13
2.3. Tests d'ajustement . . . . .	14
2.4. Détermination des paramètres par les moments . . . . .	16
2.5. Détermination des paramètres par la méthode du maximum de vraisemblance . . . . .	16
2.6. Mélange de distributions.. . . .	17
2.7. Cas de troncature. . . . .	18
III. — Distribution continue uniforme. . . . .	19
IV. — Distribution de GAUSS ou distribution normale . . . . .	22
V. — Distribution de GUMBEL ou doublement exponentielle . . . . .	27
VI. — Distribution Gausso-logarithmique . . . . .	30
VII. — Distribution GAMMA incomplète . . . . .	38
VIII. — Distribution exponentielle généralisée . . . . .	49
IX. — Distribution BETA incomplète . . . . .	58
X. — Choix d'une loi. . . . .	65
XI. — Inversion de fonctions . . . . .	68
XII. — Programmes FORTRAN . . . . .	73

# I. — GÉNÉRALITÉS

## 1.1. — Définitions.

Les fonctions de répartition étudiées ci-dessous ne concernent que des « variables aléatoires » (ou « variates ») continues, c'est-à-dire des variates pouvant prendre toutes les valeurs de l'ensemble des nombres réels dans l'intervalle  $-\infty, +\infty$ . La loi de probabilité de la variate  $X$  sera définie par la connaissance de la probabilité  $F(x)$  — au non dépassement — pour que soient vérifiées les inégalités.

$$-\infty \leq X < x$$

La première condition posée est que la « fonction de répartition »  $F(x)$  soit monotone et croissante de 0 à 1 lorsque  $x$  parcourt l'intervalle  $x_a, x_b$ , « intervalle de définition de la variate »,

$$\text{tel que : } -\infty \leq x_a < x_b \leq +\infty$$

$$\text{avec, pour : } x \leq x_a, \quad F(x) = 0,$$

$$\text{pour : } x_b \leq x, \quad F(x) = 1.$$

Nous n'étudierons que des lois pour lesquelles la fonction de densité  $f(x)$ , dérivée de  $F(x)$  par rapport à  $x$ , est continue (et forcément positive) dans l'intervalle de définition de la variate, qui ne peut prendre des valeurs nulles ou infinies qu'aux bornes de l'intervalle de définition, et dont la dérivée  $f'(x)$  par rapport à  $x$  ne s'annule, au plus, qu'une seule fois entre les bornes (fonction de densité unimodale ou amodale).

Les variations de  $F(x)$  et de  $f(x)$  peuvent se résumer comme suit :

$$-\infty \leq x \leq x_a \quad F(x) = 0 \quad f(x) = 0,$$

$$x_a < x < x_b \quad F(x) = \int_{x_a}^x f(x) dx \quad f(x) > 0,$$

$$x_b \leq x \leq +\infty \quad F(x) = 1 \quad f(x) = 0.$$

## 1.2. — Paramètres.

La fonction de répartition — ou loi de probabilité — est définie d'une part par sa formule mathématique, toujours choisie a priori par les hydrologues d'après leurs idées préconçues ou des habitudes traditionnelles, d'autre part par les valeurs numériques des paramètres qui rentrent dans l'expression analytique, valeurs qui seront estimées grâce à l'échantillon des observations dont on dispose.

Ces paramètres ont des fonctions très différentes :

- Paramètres de forme : pouvant être absent (lois de Gauss, de Gumbel), unique (loi gauss-log, gamma incomplète, exponentielle généralisée) ou multiples (deux dans la loi bêta incomplète).
- Paramètres de position et d'échelle : l'écriture analytique la plus simple de la fonction de répartition correspond à la variate « réduite » (sans dimensions) :

$$u = \frac{x - x_0}{s},$$

$x_0$  et  $s$  exprimés dans les mêmes dimensions que la variate observée  $x$ .

- Le paramètre d'échelle «  $s$  » est positif si  $F$  croît quand  $u$  croît, négatif dans le cas contraire.

— Le paramètre de position «  $x_0$  » se trouve être le mode dans les lois étudiées ci-après lorsque  $u$  varie de  $-\infty$  à  $+\infty$  (loi de Gumbel, loi de Gauss où mode et moyenne sont confondus) ; il est toujours une borne lorsque l'intervalle de définition de  $x$  est borné par une valeur finie (loi gaussio-log, gamma incomplète, exponentielle généralisée) : une des bornes de l'intervalle de définition de  $u$  a donc alors la valeur zéro.

Si l'intervalle de définition de  $x$  est compris entre deux bornes finies  $x_0$  et  $x_1$  (loi bêta incomplète) nous pouvons admettre, soit qu'il y ait deux paramètres de position et aucun paramètre d'échelle, soit qu'il y ait un paramètre d'échelle ( $x_1 - x_0$ ) et un paramètre de position  $x_0$  (ou  $x_1$ ). Les bornes de l'intervalle de définition de  $u$  sont alors zéro et un.

Les valeurs numériques de certains de ces paramètres peuvent être connues ou choisies *a priori*, sinon elles seront calculées d'après les valeurs prises par la variate dans l'échantillon observé.

### 1.3. — Valeurs centrales, moments.

Dans ce paragraphe, nous considérons une population-mère parfaitement connue (expression mathématique de la fonction de répartition et valeurs numériques « vraies » des paramètres). La variate  $X$  est supposée comprise dans l'intervalle  $x_0, x_1$  ( $x_0 < x_1$ ,  $x_0$  et /ou  $x_1$  pouvant être infinis en valeurs absolues).

#### 1.3.1. — Valeurs centrales (population-mère parfaitement connue).

— La *moyenne* est définie par : 
$$\int_{x_0}^{x_1} x f(x) dx,$$

c'est la moyenne arithmétique de toutes les valeurs comprises dans la population-mère, celles-ci étant en nombre infini pour une variate continue.

— Le *mode*  $m$  est défini par :  $f'(m) = 0$   $x_0 < m < x_1$ , c'est la valeur de la variate pour laquelle la densité de probabilité est maximale. Si  $f'(x)$  ne s'annule pas dans l'intervalle  $x_0, x_1$  (fonction de densité amodale), on peut avoir un mode observable en  $x_0$  ou en  $x_1$ , aux bornes de l'intervalle de définition de la variate  $X$ . S'il existe des modes observables en  $x_0$  et en  $x_1$  (lois bêta incomplètes par exemple, pour certaines valeurs des paramètres de forme),  $f'(x) = 0$ , définit alors la valeur de la variate pour laquelle la densité de probabilité est minimale.

— La *médiane* est définie par : 
$$\int_{x_0}^x f(x) dx = 1/2,$$

ou :

$$\int_{x_0}^x f(x) dx = \int_x^{x_1} f(x) dx,$$

une sur deux de toutes les valeurs comprises dans la population-mère est inférieure à la médiane.

Dans la loi normale, ces trois valeurs centrales sont confondues. Dans les lois unimodales on les trouve — en général — dans l'ordre (croissant ou décroissant) mode — médiane — moyenne et liées par la relation approximative :

$$(\text{moyenne} - \text{médiane}) \approx 1/3 (\text{moyenne} - \text{mode}).$$

Lorsque la moyenne est supérieure à la médiane et au mode, la dissymétrie est « positive » et la distribution est « étalée sur la droite ». Dans le cas contraire, la dissymétrie est « négative » et la distribution est « étalée sur la gauche ».

Ces trois valeurs ont des rôles différents et importants : en pensant en termes de « jeu », on peut voir que, suivant les conditions du pari, il peut être préférable de miser sur la moyenne, sur la médiane ou sur le mode.

1.3.2. — Moments (population-mère parfaitement connue).

1.3.2.1. — Le moment d'ordre  $r$  ( $r$  entier) est défini par :

$$m_r = \int_{x_0}^{x_1} x^r f(x) dx,$$

Les moments d'ordre négatif n'ont de sens que si l'intervalle de définition de  $x$  admet une borne inférieure finie, et à condition que pour  $x = 0$  l'expression  $x^r f(x)$  soit identiquement nulle.

1.3.2.2. — En notant  $\bar{x}$  la valeur de la moyenne, on définit le moment centré d'ordre  $r$  par l'expression ( $r$  entier) :

$$\mu_r = \int_{x_0}^{x_1} (x - \bar{x})^r f(x) dx,$$

Seuls, les moments centrés d'ordre positif ont un sens, dans la mesure où les moments de même ordre et d'ordres inférieurs existent.

Le moment centré d'ordre 1 est nul.

Le moment centré d'ordre 2 porte un nom particulier : c'est la variance  $\mu_2 = m_2 - m_1^2$ .

Les moments centrés d'ordre  $r$  se déduisent facilement des moments non centrés en considérant le développement de  $(x - \bar{x})^r$

$$\mu_r = m_r + (-1) \frac{r!}{1!(r-1)!} \bar{x} m_{r-1} + \dots + (-1)^i \frac{r!}{i!(r-i)!} \bar{x}^i m_{r-i} + \dots + (-1)^r \bar{x}^r,$$

d'où :

$$\mu_3 = m_3 - 3\bar{x} m_2 + 2\bar{x}^3$$

$$\mu_4 = m_4 - 4\bar{x} m_3 + 6\bar{x}^2 m_2 - 3\bar{x}^4.$$

Inversement, on a la relation (où  $\mu_1 = 0$ )

$$m_r = \mu_r + \frac{r!}{(r-1)!} \bar{x} \mu_{r-1} + \dots + \frac{r!}{i!(r-i)!} \bar{x}^i \mu_{r-i} + \dots + \bar{x}^r.$$

d'où :

$$m_3 = \mu_3 + 3\bar{x} \mu_2 + \bar{x}^3,$$

$$m_4 = \mu_4 + 4\bar{x} \mu_3 + 6\bar{x}^2 \mu_2 + \bar{x}^4.$$

1.3.2.3. — La variance d'un moment (centré ou non) est donnée par l'expression générale :

$$\text{Var } m_r = m_{2r} - m_r^2.$$

La covariance de deux moments (pris par rapport à la même origine) est donnée par :

$$\text{Covar } (m_r, m_i) = m_{r+i} - m_r m_i.$$

La variance d'une fonction linéaire de moments :

$\varphi(m_1 \dots m_i \dots m_2)$  s'exprime par :

$$\text{Var } \varphi = \Sigma \left( \frac{\partial \varphi}{\partial m_i} \right)^2 \text{Var } m_i + 2 \Sigma \left( \frac{\partial \varphi}{\partial m_i} \right) \left( \frac{\partial \varphi}{\partial m_k} \right) \text{Covar } (m_k, m_i) \quad [k \neq i].$$

1.3.3. — Cumulants (population-mère parfaitement connue).

Sans nous étendre sur les fonctions caractéristiques, nous allons signaler l'existence des cumulants.

Dans la mesure où les moments d'ordre positif existent, les cumulants  $K_r$  (cumulants d'ordre  $r$ ,  $r$  entier positif) se déduisent de ces moments en égalant les coefficients des mêmes puissances de  $t$  dans l'expression :

$$1 + m_1 \frac{t}{1!} + m_2 \frac{t^2}{2!} + \dots + m_r \frac{t^r}{r!} + \dots = \left[ 1 + K_1 \frac{t}{1!} + K_2 \frac{t^2}{2!} + \dots \right] \\ \left[ 1 + K_2 \frac{t^2}{2!} + \frac{K_2^2}{2!} \left( \frac{t^2}{2!} \right)^2 + \dots \right] \\ \left[ 1 + K_r \frac{t^r}{r!} + \frac{K_r^2}{2!} \left( \frac{t^r}{r!} \right)^2 + \dots + \frac{K_r^i}{i!} \left( \frac{t^r}{r!} \right)^i + \dots \right]$$

On peut les déduire aussi des moments centrés en égalant les coefficients des mêmes puissances de  $t$  dans l'expression :

$$K_1 t + K_2 \frac{t^2}{2!} + \dots + K_r \frac{t^r}{r!} + \dots = m_1 t + U - \frac{1}{2} U^2 + \frac{1}{3} U^3 + \dots + (-1)^i \frac{1}{i!} U^i + \dots$$

où :

$$U = \frac{\mu_2}{2!} t^2 + \frac{\mu_3}{3!} t^3 + \dots + \frac{\mu_i}{i!} t^i + \dots$$

Les premiers cumulants s'écrivent donc :

$$K_1 = m_1 = \bar{x},$$

$$K_2 = \mu_2 = m_2 - m_1^2,$$

$$K_3 = \mu_3 = m_3 - 3 m_1 m_2 + 2 m_1^3,$$

$$K_4 = \mu_4 - 3 \mu_2^2 = m_4 - 4 m_1 m_3 - 3 m_2^2 + 12 m_1^2 m_2 - 6 m_1^4.$$

Comme les moments centrés, les cumulants, *sauf le premier*, ne dépendent pas de l'abscisse-origine de la variate. Le moment centré d'ordre  $r$ , le cumulant d'ordre  $r$  ( $r \neq 1$ ) se déduisent respectivement du moment centré d'ordre  $r$  ou du cumulant d'ordre  $r$  de la variate réduite par simple multiplication par  $s^r$ ,  $s$  étant le paramètre d'échelle.

L'intérêt des cumulants provient de ce que, lorsque l'expression d'une variate est une combinaison linéaire de plusieurs variates indépendantes :  $X = \alpha_1 X_1 + \dots + \alpha_i X_i$ , chaque cumulant de  $X$  est la somme des cumulants de même ordre des variables indépendantes, pondérés comme suit :

$$K_r(X) = \alpha_1^r K_r(X_1) + \dots + \alpha_i^r K_r(X_i).$$

#### 1.3.4. — Coefficients d'asymétrie et d'aplatissement (population-mère parfaitement connue).

Coefficient d'asymétrie :  $\gamma_1 = K_3/K_2^{3/2} = \mu_3/\mu_2^{3/2}$ .

si  $\gamma_1$  est positif la distribution est étalée sur la droite, s'il est négatif elle est étalée sur la gauche. Si  $\gamma_1$  est nul la distribution n'est pas forcément symétrique (par exemple lois exponentielles généralisées), mais si la distribution est symétrique le coefficient  $\gamma_1$  est nul. Le champ de variation de  $\gamma_1$  n'est pas borné par des valeurs finies, positive ou négative.

Coefficient d'aplatissement :  $\gamma_2 = K_4/K_2^2 = \frac{\mu_4}{\mu_2^2} - 3$ .

si  $\gamma_2$  est positif la distribution est moins aplatie que la distribution « normale ». S'il est négatif elle est plus aplatie. Le champ de variation de  $\gamma_2$  est borné inférieurement par la valeur  $-2$  et n'a pas de borne supérieure.

1.3.5. — Moyennes harmonique et géométrique (population-mère parfaitement connue).

Moyenne harmonique : c'est l'inverse du moment d'ordre  $-1$  dont la valeur moyenne est :

$$\int_{x_0}^{x_1} \frac{1}{x} f(x) dx, \text{ qui n'a de sens que si la variate est bornée par une valeur finie } x_0 \geq 0$$

(ou  $x_1 \leq 0$ ) et si la densité de probabilité est identiquement nulle pour  $x = 0$ .

La variance de l'inverse de la moyenne harmonique est :  $m_{-2} - (m_{-1})^2$  à condition que  $m_{-2}$  existe.

Moyenne géométrique : c'est l'exponentielle de la valeur moyenne du logarithme naturel de la variate, qui est  $\int_{x_0}^{x_1} \text{Log}(x) f(x) dx$  et n'a de sens que pour  $x_0 > 0$  (mais il peut y avoir d'autres conditions d'existence dépendant de  $f(x)$ ).

La variance de la valeur moyenne du logarithme est :

$$\int_{x_1}^{x_0} \text{Log}^2(x) f(x) dx - \left[ \int_{x_0}^{x_1} \text{Log}(x) f(x) dx \right]^2.$$

Lorsque les moyennes harmonique et géométrique existent, elles sont dans l'ordre :

zéro < moyenne harmonique < moyenne géométrique < moyenne arithmétique.

#### 1.4. — Loïs tronquées et troncatures.

##### 1.4.1. — Loïs tronquées.

1.4.1.1. — Nous n'avons jusqu'ici envisagé que les cas où pour  $x \leq a$  nous avons  $F(x) = 0$ . Dans bien des cas, on est conduit par l'observation à utiliser des loïs tronquées en fréquence, suivant le schéma :

$$\begin{aligned} x \leq a & \quad F(x) = F_0, \\ a < x < b & \quad F(x) = \int_a^x f(x) dx, \\ b \leq x & \quad F(x) = F_1 \quad (\text{Figure 1}). \end{aligned}$$

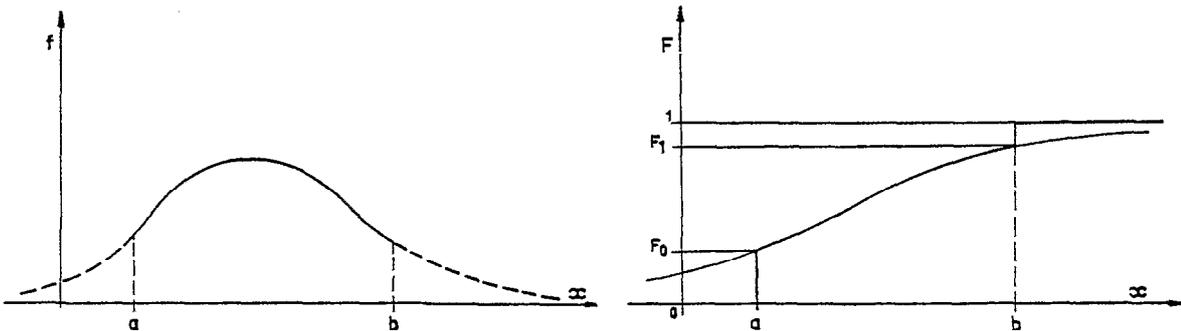


FIG. 1.

1.4.1.2. — Un cas particulier, qui simplifie bien des problèmes, est celui d'une variate dont l'intervalle de définition admet  $a = x_0$  comme borne inférieure de valeur finie (c'est-à-dire

borne de valeur nulle pour la variable réduite  $u$ ) la borne supérieure de l'intervalle de définition étant  $b = x_1$  qui peut être infini (Figure 2).

Si  $\Phi$  est la fonction de répartition de la variable réduite  $u$ , la fonction de répartition de  $x$  s'écrira :

$$\begin{aligned} x \leq x_0 & \quad F(x) = F_0, \\ x_0 < x < x_1 & \quad F(x) = F_0 + (F_1 - F_0) \Phi, \\ x_1 \leq x & \quad F(x) = F_1. \end{aligned}$$

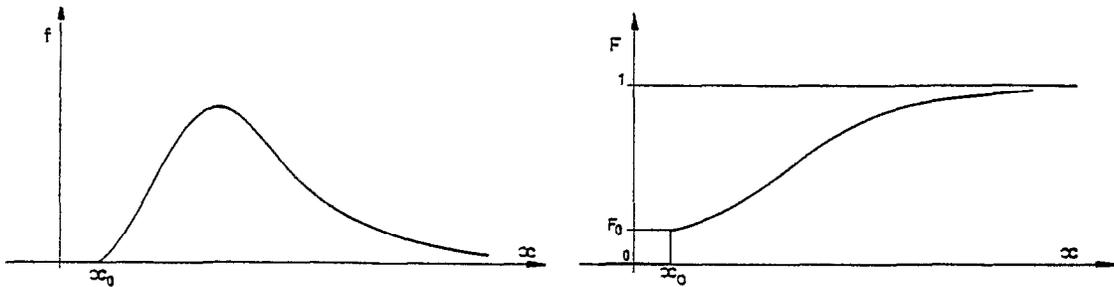


FIG. 2.

En hydrologie, on n'a envisagé jusqu'ici que des lois tronquées vers les fréquences basses :  $F_1 = 1$ . Il n'est pas impossible de faire un tronquage vers les fréquences hautes, ou des deux bords.

#### 1.4.2. — Troncature des observations.

1.4.2.1. — On peut être obligé de considérer des troncutures des observations, par exemple lorsque la densité de probabilité ne tend pas vers zéro lorsque  $x$  tend vers une borne de son intervalle de définition, ce qui se produit en général lorsqu'on utilise une loi tronquée en fréquence. On ne peut plus distinguer entre les valeurs  $x = x_0$  et les valeurs  $x = x_0 + \varepsilon$ . C'est le cas, entre autres, des observations pluviométriques journalières où ne sont notées que les hauteurs supérieures à 0,1 mm ; nous connaissons le nombre des observations supérieures ou égales à  $x_0 + \varepsilon$ , ainsi que les valeurs correspondantes de la variate, mais non le nombre des observations qui auraient dû être faites entre  $x_0$  et  $x_0 + \varepsilon$ , ni leurs valeurs numériques :  $x_h = x_0 + \varepsilon$  est le seuil de troncuture. (Figure 3).

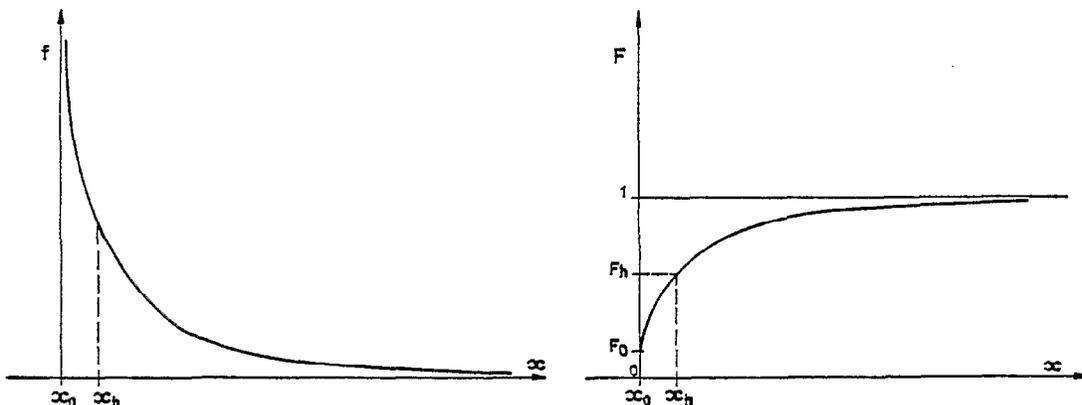


FIG. 3.

Bien entendu, ce premier genre de troncature peut être fait du côté de la borne inférieure, du côté de la borne supérieure ou des deux côtés de l'intervalle de définition de  $x$ .

1.4.2.2. — Un autre cas de troncature un peu différent est celui, par exemple, des observations pluviométriques de hauteurs fortes, supérieures à la capacité  $\eta$  du pluviomètre : nous connaissons le nombre des observations supérieures à  $\eta$  mais non leurs valeurs numériques.

### 1.5. — Moments déduits d'un échantillon de taille connue.

Dans le paragraphe 1.3 nous avons examiné les valeurs centrales et les moments déduits de la loi de répartition supposée complètement connue d'une population-mère. Dans la réalité, on dispose d'un échantillon d'observations que l'on suppose indépendantes et provenant d'une seule population-mère par des tirages non exhaustifs. Dans ce paragraphe 1.5 nous ne considérons que :  
 — les cas où il n'y a pas de troncature : c'est-à-dire que nous connaissons la taille  $n$  de l'échantillon (nombre d'observations) et la valeur numérique de chaque observation  $x_1 x_2 \dots x_i \dots x_n$  ;  
 — les cas où la loi n'est pas tronquée en fréquence.

1.5.1. — La *moyenne* arithmétique est calculée par :  $m_1 = \frac{1}{n} \sum_1^n x_i$ , le moment d'ordre  $r$  est

calculé par  $m_r = \frac{1}{n} \sum_1^n x_i^r$ , entre autres la *moyenne harmonique*  $m_h$  est calculée par :

$\frac{1}{m_h} = \frac{1}{n} \sum_1^n \frac{1}{x_i}$ , la *moyenne géométrique*  $m_g$  est calculée par  $\text{Log}(m_g) = \frac{1}{n} \sum_1^n \text{Log}(x_i)$ .

Il se peut donc que des expressions, qui n'ont pas de sens en analyse intégrale d'après la forme mathématique de la loi choisie pour représenter la répartition de la population-mère, conduisent à des valeurs numériques compréhensibles dans le cas d'un échantillon de dimension finie.

1.5.2. — La *variance* (cumulant ou moment centré d'ordre 2) est calculée par :

$$K_2 = \mu_2 = \frac{1}{n-1} \left[ \sum_1^n x_i^2 - \frac{1}{n} \left( \sum_1^n x_i \right)^2 \right].$$

de même la variance du moment d'ordre  $r$  est calculée par :

$$\text{Var } m_r = \frac{1}{n-1} \left[ \sum_1^n x_i^{2r} - \frac{1}{n} \left( \sum_1^n x_i^r \right)^2 \right].$$

Le cumulant (ou moment centré) d'ordre 3 est calculé par :

$$K_3 = \mu_3 = \frac{1}{(n-1)(n-2)} \left[ n \sum_1^n x_i^3 - 3 \sum_1^n x_i \sum_1^n x_i^2 + \frac{2}{n} \left( \sum_1^n x_i \right)^3 \right].$$

Le cumulant d'ordre 4 est calculé par :

$$K_4 = \frac{1}{(n-1)(n-2)(n-3)} \left[ (n^2 + n) \sum_1^n x_i^4 - 4(n+1) \sum_1^n x_i \sum_1^n x_i^3 - 3(n-1) \left( \sum_1^n x_i^2 \right)^2 + 12 \left( \sum_1^n x_i \right)^2 \sum_1^n x_i^2 - \frac{6}{n} \left( \sum_1^n x_i \right)^4 \right] = \frac{n^2}{(n-1)(n-2)(n-3)} \left[ (n+1) \mu_4 - 3(n-1) \mu_2^2 \right].$$

1.5.3. — Le théorème central limite peut s'énoncer.

Si  $Z$  est une combinaison linéaire de  $n$  variates  $X$  indépendantes, quelle que soit la loi suivie par chacun des  $X$ , la loi de répartition de  $Z$  tend vers une loi normale lorsque  $n$  augmente indéfiniment, à condition que l'ordre de grandeur de certaines des variates ne domine pas trop nettement celui des autres. La convergence est plus rapide si chacun des  $X$  suit une loi unimodale, et d'autant plus rapide que ces lois unimodales sont moins dissymétriques.

Si l'on est placé dans les conditions d'application de ce théorème : la variance d'une fonction — même non linéaire — de moments  $\varphi(m_1 \dots m_i \dots m_n)$  pourra approximativement être estimée par :

$$\text{Var } \varphi = \sum \left( \frac{\partial \varphi}{\partial m_i} \right)^2 \text{Var } m_i + 2 \sum \left( \frac{\partial \varphi}{\partial m_i} \right) \left( \frac{\partial \varphi}{\partial m_k} \right) \text{Covar } (m_i, m_k) \quad [k \neq i].$$

1.6. — Mélange de distributions.

Bornons-nous au cas de deux distributions ; la généralisation est facile. Si  $G_1$  et  $G_2$  sont les fonctions de répartition des deux populations-mères, mélangées dans les proportions  $p$  et  $(1 - p)$  et de fonctions de densité respectives  $g_1$  et  $g_2$ , la fonction de répartition de la population résultante sera :

$$F(x) = p G_1(x) + (1 - p) G_2(x),$$

et la fonction de densité :

$$f(x) = p g_1(x) + (1 - p) g_2(x),$$

pouvant être plurimodale ;

d'où le moment — non centré — d'ordre  $r$  :

$$m_r = p m_{1r} + (1 - p) m_{2r}.$$

## II. — DÉTERMINATION DES PARAMÈTRES

2.1. — Nous nous trouvons en présence d'un échantillon de  $n$  observations, supposées indépendantes, et nous avons choisi la forme mathématique de la loi de probabilité devant représenter la loi de répartition de la population-mère de laquelle a été extrait l'échantillon. Nous avons à calculer le mieux possible les valeurs numériques des paramètres, en tenant compte au maximum des informations enfermées dans l'échantillon.

Les trois principales méthodes utilisées procèdent :

- par le maximum de vraisemblance ;
- par les moments de l'échantillon ;
- par minimisation d'un test d'adéquation.

### 2.2. — Qualificatifs des estimations.

Soit un paramètre de valeur inconnue  $\theta$  à déterminer, un échantillon de  $n$  valeurs indépendantes de la variate  $X$  et l'estimation du paramètre  $\theta$  par :

$$\theta' = \varphi(x_1, x_2 \dots x_n).$$

Si, à chaque série possible de  $n$  épreuves, on associe la valeur correspondante de  $\theta'$ , cette valeur peut être considérée comme une valeur possible d'une variable aléatoire  $Y$ .

#### 2.2.1. — Estimation *consistante* ou *correcte*.

$\theta'$  est une estimation consistante de  $\theta$  si la variate  $Y$  vérifie les conditions suivantes :

Lorsque  $n$  croît indéfiniment :

- la valeur limite de l'espérance mathématique de  $Y$  est  $\theta$  (c.a.d. la valeur moyenne de  $Y$  tend vers  $\theta$ ) ;
- la valeur limite de la variance de  $Y$  est nulle.

L'estimation de la valeur du paramètre est d'autant plus correcte que la taille de l'échantillon est plus grande. Par exemple, dans une distribution normale la valeur centrale peut être déterminée par la moyenne ou la médiane observée.

#### 2.2.2. — Estimation *absolument correcte*.

$\theta'$  est une estimation absolument correcte de  $\theta$  si la variate  $Y$  vérifie les conditions suivantes :

- quel que soit  $n$ , l'espérance mathématique de  $Y$  est  $\theta$  ;
- si  $n$  croît indéfiniment, la valeur limite de la variance de  $Y$  est nulle.

Par exemple, l'estimation absolument correcte d'une variance, lorsque la valeur moyenne de la population-mère est estimée d'après l'échantillon, est :  $\frac{1}{n-1} \left[ \sum x_i^2 - \frac{1}{n} (\sum x_i)^2 \right]$ , alors qu'une estimation consistante est :  $\frac{1}{n} \left[ \sum x_i^2 - n \bar{x}^2 \right]$ .

### 2.2.3. — Estimation exhaustive ou *efficace*.

C'est celle qui utilise la totalité de l'information disponible dans l'échantillon. Elle correspond à l'estimation du paramètre affectée de la variance minimale possible pour une taille donnée de l'échantillon. Dans l'exemple d'une distribution normale, la moyenne est une estimation efficace de la valeur centrale mais la médiane n'en est pas une estimation efficace.

2.2.4. — D'après Borel, la méthode du maximum de vraisemblance donne toujours des estimations correctes. Dans le cas d'un seul paramètre à estimer, et pour un nombre infiniment grand d'observations, l'estimation qu'elle donne est celle qui comporte la plus petite variance.

D'après Fischer, il est démontré, dans la théorie des grandes séries, qu'aucune estimation de paramètre ne peut avoir une variance aléatoire inférieure à celle qui est fournie par la méthode du maximum de vraisemblance. Le groupe des estimations — auquel appartient aussi dans tous les cas la solution du minimum de  $\chi^2$  — qui s'accordent dans leurs variances aléatoires avec la solution du maximum de vraisemblance, possède une importance particulière. On les désigne sous le nom d'estimations efficaces.

### 2.3. — Tests d'ajustement.

Nous pensons nécessaire d'insister sur les tests d'adéquation : grâce au calcul par ordinateur, ils seront de plus en plus utilisés. Ces tests d'adéquation peuvent répondre à deux questions différentes.

#### 2.3.1. — Cas « A ».

Une loi de distribution étant connue (par sa forme mathématique et les valeurs numériques des paramètres) et décrivant de façon parfaite une population-mère, quelle est la probabilité pour qu'un échantillon donné puisse être considéré comme étant tiré de la population-mère.

#### 2.3.2. — Cas « B ».

Une loi de distribution étant définie par sa forme mathématique choisie *a priori* et par les valeurs numériques des paramètres estimés d'après un échantillon donné, quelle est la probabilité pour que la loi de distribution représente effectivement la population-mère dont l'échantillon est représentatif *a priori*.

#### 2.3.3. — Niveau de signification et puissance d'un test.

Les notions ci-dessous sont générales et non spécifiques aux tests d'adéquation.

L'hypothèse testée peut être vraie ou fausse. Il y a deux erreurs possibles :

- risque de première espèce : « V » on peut rejeter à tort l'hypothèse si elle est vraie ;
- risque de seconde espèce : « F » on peut l'accepter à tort si elle est fausse.

La probabilité de « V » est le niveau de signification «  $\alpha$  » défini par la valeur numérique du test ( $\alpha$  est toujours donné en probabilité au dépassement de cette valeur numérique).

La probabilité de « F » est «  $\beta$  ». La puissance du test est  $(1 - \beta)$ , qui dépend de la formulation du test, et décroît avec la taille de l'échantillon et le niveau de signification  $\alpha$ .

Avec de petits échantillons et le niveau de signification habituellement et arbitrairement utilisé de 5 %, le risque de deuxième espèce devient très grand. Dans l'application de certains tests (par exemple comparaison de moyennes...) il est possible d'harmoniser les deux risques ou de les fixer à des seuils différents : entre autres, contrôle sur échantillon de la qualité d'une série de pièces avec distinction du risque « client » et du risque « producteur ».

En fait, en ce qui concerne les tests d'adéquation, leur puissance n'est pas calculable : on ne peut que comparer les tests entre eux pour savoir lequel sera le plus puissant.

### 2.3.4. — Qualités d'un test d'ajustement.

Les principales qualités que nous lui demandons sont :

- utiliser au mieux, avec la plus grande puissance possible du test, toutes les valeurs de l'échantillon dont on dispose, et avec le même poids pour chaque élément de l'échantillon ;
- tenir compte du fait que l'hydrologue non seulement choisit *a priori* la forme mathématique de la loi de répartition, mais encore calcule les valeurs des paramètres de cette loi d'après l'échantillon dont il dispose (cas « B » paragraphe 2.3.2.).

### 2.3.5. — Test du $\chi^2$ .

Le seul test que nous connaissons à posséder cette dernière qualité est le test du  $\chi^2$  ; il ne possède pas la première, et nous pensons qu'il serait possible de construire un test qui l'aurait et serait plus puissant. A propos du test de  $\chi^2$  nous ferons les remarques suivantes :

La plupart des auteurs demandent de prendre comme nombre de degrés de liberté du  $\chi^2$  le nombre de classes moins le nombre de liaisons ayant servi à l'ajustement des paramètres ou à l'application du test. Ils estiment, après Fischer, que cette diminution du nombre de degrés de liberté permet de tester l'adéquation de la loi choisie (avec les paramètres calculés) comme représentation de l'échantillon et de la population-mère. Comme liaison servant à l'application du test de  $\chi^2$ , il faut expliciter, dans le nombre de degrés de liberté à retrancher, la condition (si on l'utilise) d'égalité des effectifs de l'échantillon observé et de l'échantillon « théorique » de comparaison. Dans les tests de Cramer, Anderson et même Kolmogorov (qui ne correspondent malheureusement qu'au cas « A » paragraphe 2.3.1.) cette liaison est implicitement prise en compte, mais ces tests ne donnent pas les moyens de prendre en compte les autres liaisons explicitées par la méthode de calcul des paramètres (par vraisemblance maximale ou moments ou minimisation du test).

Le test du  $\chi^2$ , par suite du groupement des observations en classes, n'a pas la première qualité demandée plus haut. Pour des échantillons de mêmes effectifs, le test du  $\chi^2$  est beaucoup moins puissant qu'un test utilisant chaque élément de l'échantillon indépendamment des autres : par exemple (nous sommes dans le cas « A ») si  $n$  est l'effectif nécessaire pour garantir une certaine puissance du test de Cramer, l'effectif nécessaire pour garantir la même puissance, au même niveau de signification, du test de  $\chi^2$  sera de l'ordre de  $n^{5/4}$  (passant, pour  $n = 80$ , à 240 et, pour  $n = 40$ , à 100).

Un autre inconvénient du test du  $\chi^2$  est que la valeur numérique du test dépend trop du mode de découpage en classes, découpage qui est laissé, en principe, à la discrétion du calculateur. Nous donnerons comme exemple l'échantillon de 45 totaux pluviométriques d'années consécutives à Ziguinchor. Loi choisie : « gamma incomplète » à deux paramètres calculés par les moments avec amélioration. Le découpage de l'échantillon en 7 classes peut conduire de  $\chi^2 = 0$  probabilité 1 (au dépassement) à  $\chi^2 = 14,25$  probabilité 0,005 ; en découpant en classes d'égales probabilités théoriques, l'on trouve :

Pour 9 classes $\chi^2 = 8,40$	avec probabilité de dépassement = 0,21
Pour 8 classes $\chi^2 = 8,19$	avec probabilité de dépassement = 0,15
Pour 7 classes $\chi^2 = 0,578$	avec probabilité de dépassement = 0,96
Pour 6 classes $\chi^2 = 2,34$	avec probabilité de dépassement = 0,50
Pour 5 classes $\chi^2 = 3,56$	avec probabilité de dépassement = 0,18
Pour 4 classes $\chi^2 = 1,31$	avec probabilité de dépassement = 0,25

Le fait de découper l'échantillon en créant des classes d'au moins cinq unités « théoriques » aux deux extrémités du rangement notamment, restreint terriblement la valeur du test du  $\chi^2$  qui renseigne seulement sur la possibilité qu'à la loi choisie (avec ses paramètres calculés) de représenter la distribution observée dans sa zone de forte densité de probabilité. Ainsi nous pouvons admettre que la distribution des pluviométries annuelles à Ziguinchor est convenablement repré-

sentée par une loi gamma incomplète du rang 5 au rang 40 (pluviométries annuelles rangées en ordre croissant et découpage en 9 classes d'égalles probabilités) c'est-à-dire pour des temps de récurrence inférieurs à 9 ans, car le groupement en classes escamote les distorsions entre l'échantillon et la loi ajustée, et surtout dans le cas des valeurs extrêmes.

2.3.6. — En conclusion de cet article 2.3, nous pensons que la minimisation d'un test d'ajustement n'est guère utilisable : le seul test possible — à l'heure actuelle — étant celui du  $\chi^2$  que nous ne trouvons ni assez puissant ni assez consistant.

#### 2.4. — Détermination des paramètres par les moments.

C'est une méthode relativement simple à appliquer lorsqu'on n'a pas d'ordinateur à sa disposition.

On écrit d'après la loi choisie les équations donnant les moments en fonction des paramètres, à partir du moment d'ordre 1. On calcule, à partir des observations, autant de moments des valeurs observées qu'il y a de paramètres à déterminer. On obtient ainsi un système d'équations — souvent implicites et transcendantes — en nombre égal à celui des paramètres, système qu'il faut résoudre.

Les estimations ainsi obtenues sont consistantes mais généralement non efficaces.

Dans un cas simple, la loi gamma incomplète sans paramètre de position, nous avons pu mettre au point une amélioration de la méthode en tenant compte des covariances entre moments : l'expérience nous a montré que l'on obtenait alors des estimations aussi « correctes » qu'avec la méthode du maximum de vraisemblance.

#### 2.5. — Détermination des paramètres par la méthode du maximum de vraisemblance.

2.5.1. — Ce qui suit est directement tiré du traité de M. Roche « Hydrologie de Surface ».

Supposons qu'un échantillon, tiré d'une population-mère, comporte  $n$  valeurs  $x_i$  pouvant se produire chacune avec la probabilité  $P_i$ . La probabilité pour qu'un échantillon de  $n$  valeurs obtenues par tirages indépendants soit précisément l'échantillon observé est :

$$P_1 P_2 P_3 \dots P_i \dots P_n.$$

On appelle cette probabilité « vraisemblance de l'échantillon ».

La méthode du maximum de vraisemblance consiste à déterminer les paramètres de la loi de façon à rendre l'échantillon le plus vraisemblable possible.

Si la variate est continue, chacun des termes ci-dessus et le produit lui-même sont infiniment petits. On définit alors la vraisemblance de l'échantillon comme une quantité proportionnelle au produit des densités de probabilité, c'est-à-dire à :

$$\mathcal{L} = f_1 f_2 f_3 \dots f_n,$$

avec :

$$f_i = f(x_i, a, b \dots k),$$

$a, b \dots k$  paramètres de la loi de probabilité dont les valeurs sont inconnues et indépendantes les unes des autres.

Le but cherché est de maximiser  $\mathcal{L}$ , donc d'annuler les dérivées partielles de  $\mathcal{L}$  par rapport aux différents paramètres, ce qui donne le système d'équations :

$$\frac{\partial \mathcal{L}}{\partial a} = 0 \quad \frac{\partial \mathcal{L}}{\partial b} = 0 \quad \dots \quad \frac{\partial \mathcal{L}}{\partial k} = 0.$$

Il est plus simple d'écrire :

$$\text{Log } \mathcal{L} = \sum_1^n \text{Log } f_i,$$

et le système ci-dessus est remplacé par :

$$\sum_1^n \frac{1}{f_i} \frac{\partial f_i}{\partial a} = 0 \quad \dots \quad \sum_1^n \frac{1}{f_i} \frac{\partial f_i}{\partial k} = 0.$$

La méthode du maximum de vraisemblance conduit à des estimations correctes et efficaces, mais non forcément — dans certains cas — « absolument correctes ».

2.5.2. — Malgré cette dernière restriction, nous utiliserons toujours pour l'estimation du plus grand nombre possible de paramètres la méthode de vraisemblance maximale à cause de son efficacité.

En fait, cette méthode ne peut s'appliquer à la détermination du paramètre de position lorsqu'à la borne finie de l'intervalle de définition de la variate la densité de probabilité n'est pas identiquement nulle (loi « en J » du paragraphe 1.4.2.1., loi tronquée du paragraphe 1.4.1.1.). De même, dans les lois à deux bornes finies et si la densité de probabilité à ces bornes n'est pas identiquement nulle, le paramètre de position et le paramètre d'échelle (c'est-à-dire les deux paramètres de position) ne sont pas calculables par la vraisemblance maximale.

Le système d'équations donne comme solution pour le paramètre de position correspondant à une borne finie où la densité de probabilité n'est pas nulle, la plus petite valeur observée si la borne est limite inférieure ou la plus grande valeur observée si la borne est limite supérieure.

2.5.3. — Lois tronquées.

2.5.3.1. — Cas des lois tronquées du paragraphe 1.4.1.2.

Nous avons un paramètre supplémentaire « $F_0$ » de tronquage qui s'ajoute aux paramètres de position, d'échelle et de forme. Ce paramètre de tronquage n'est jamais calculable par la méthode du maximum de vraisemblance : il disparaît du système d'équations.

2.5.3.2. — Cas des lois tronquées avec troncature (du paragraphe 1.4.2.1.).

Lorsque la densité de probabilité ne tend pas vers zéro quand la variable réduite  $u$  tend vers zéro, la véritable valeur individuelle de chaque observation inférieure au seuil de troncature est inconnue, le nombre même de ces observations est inconnu si le paramètre de tronquage n'est pas nul, mais le paramètre de position  $x_0$  est connu par avance : c'est la valeur correspondant à la densité de probabilité maximale. Le paramètre de tronquage se détermine indirectement grâce aux autres paramètres que l'on peut estimer par le maximum de vraisemblance.

2.5.3.3. — Si la densité de probabilité tend vers zéro en même temps que la variable réduite  $u$  : le paramètre de position  $x_0$  et le paramètre de tronquage  $F_0$  sont déduits directement des observations (il n'y a pas « troncature des observations ») et seuls les paramètres d'échelle et de forme sont à estimer.

2.6. — Mélange de distributions (cf paragraphe 1.6.).

La détermination des paramètres peut se faire par le maximum de vraisemblance (sous les réserves du paragraphe 2.5.2. pour les paramètres de position). Le système d'équation devient très compliqué.

La méthode des moments peut aussi être utilisée : si nous supposons le mélange de deux distributions — non tronquées — à trois paramètres chacune, soit six paramètres, plus le paramètre de proportions du mélange, il faut utiliser les sept premiers moments. Le système d'équation est compliqué et les déterminations des paramètres sont imprécises si la taille de l'échantillon n'est pas très grande, car la variance d'un moment croît très rapidement avec son ordre.

2.7. — Cas de troncature (du paragraphe 1.4.2.2.).

Comme nous ne traiterons pas de ce cas dans les études des différentes lois, nous allons indiquer comment s'écrit l'équation du maximum de vraisemblance.

Soit  $n$  le nombre connu d'observations de valeurs connues ( $x_1 \dots x_i \dots x_n$ ),  $n_d$  le nombre connu d'observations impossibles (par exemple d'observations de débordement du pluviomètre) mais que l'on sait de valeur supérieure à un seuil  $x_d$ .

$f(x)$  étant la fonction de densité de la loi de répartition choisie :

$$\mathfrak{L} = \left[ \int_{x_d}^{+\infty} f(x) dx \right]^{n_d} \prod_{i=1}^n f(x_i),$$

Si de plus on fait une troncature du type du paragraphe 1.4.2.1. en négligeant les observations inférieures à un seuil  $x_n$  de valeur supérieure à la borne inférieure  $x_0$  de l'intervalle de définition de la variable,  $F_0$  étant le paramètre de tronquage et  $n$  le nombre, connu, d'observations de valeurs connues utilisées, l'expression donnée ci-dessus pour  $\mathfrak{L}$  est à multiplier par

$$\left[ (1 - F_0) / (1 - F_0) \int_{x_n}^{+\infty} f(x) dx \right]^{n + n_d}.$$

ce qui montre que  $F_0$  est toujours éliminé par l'application de cette méthode.

### III. — DISTRIBUTION CONTINUE UNIFORME

Nous n'en parlons que pour mémoire.

3.1. — C'est la fonction de répartition de la probabilité, quelle que soit l'expression mathématique de cette probabilité, à seule condition qu'elle soit absolument continue. Soit :  $u = F(x)$  cette probabilité — au non dépassement ou au dépassement — la variate  $u$  est bornée inférieurement par la valeur zéro et supérieurement par la valeur 1.

La distribution continue uniforme s'écrit :

$$\begin{aligned} G(u) &= 0 && \text{pour } u \leq 0, \\ G(u) &= u && \text{pour } 0 < u < 1, \\ G(u) &= 1 && \text{pour } 1 \leq u. \end{aligned}$$

3.2. — La fonction de densité est  $g(u) = 1$  pour  $0 \leq u \leq 1$  et est nulle en dehors de cet intervalle (cf figure 4).

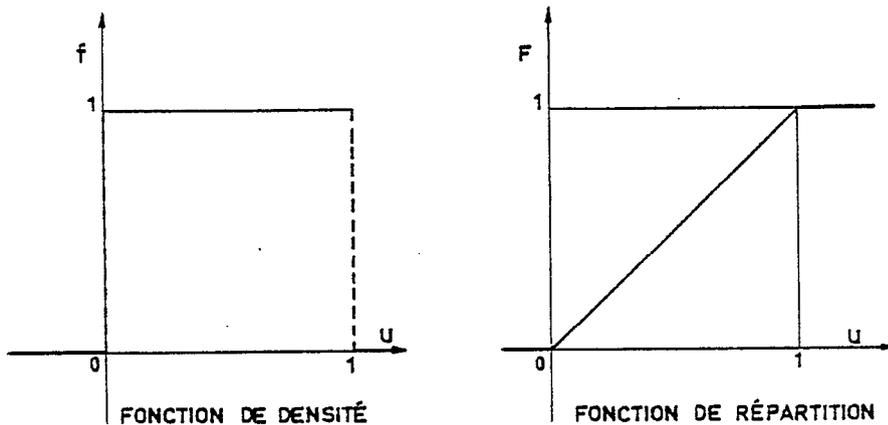


FIG. 4.

3.3. — Il n'y a pas de mode.

La distribution continue uniforme est symétrique par rapport à  $u = 1/2$ , valeur de la moyenne et de la médiane.

Les moments non centrés d'ordre  $r$  sont :  $\frac{1}{r+1}$  ( $r > 0$ ).

Les moments centrés d'ordre impair sont nuls et ceux d'ordre pair " $2r$ " sont :  $\frac{1}{(2r+1)4^r}$   
d'où :  $K_1 = 1/2$ ,  $K_2 = 1/12$ ,  $K_3 = 0$ ,  $K_4 = 1/120$ .

et les coefficients d'asymétrie  $\gamma_1 = 0$ , d'aplatissement  $\gamma_2 = -6/5$ .

3.4. — La moyenne harmonique n'existe pas.

La moyenne logarithmique est  $-1$ , d'où la moyenne géométrique  $1/e$ .

3.5. — La fonction de densité de la moyenne  $\bar{u}$  d'un échantillon de  $n$  variables indépendantes tirées d'une loi continue uniforme

est 
$$\frac{n^n}{(n-1)!} \sum_{i=0}^k (-1)^i \frac{n!}{i!(n-i)!} \left(\bar{u} - \frac{i}{n}\right)^{n-1}.$$

pour une valeur de  $\bar{u}$  :  $\frac{k}{n} \leq \bar{u} \leq \frac{k+1}{n}$  et pour tout  $k$  entier de 0 à  $n-1$ .

Par exemple, pour  $n = 2$  et

pour :  $0 \leq \bar{u} \leq 1/2$  ; la fonction de densité de  $\bar{u}$  est  $4\bar{u}$ ,

pour :  $1/2 \leq \bar{u} \leq 1$  ; la fonction de densité de  $\bar{u}$  est  $4(1-\bar{u})$ ,

pour  $n = 3$  et

pour :  $0 \leq \bar{u} \leq 1/3$  ; la fonction de densité de  $\bar{u}$  est  $\frac{27}{2}\bar{u}^2$ ,

pour :  $1/3 \leq \bar{u} \leq 2/3$  ; la fonction de densité de  $\bar{u}$  est  $\frac{27}{2} \left[ \bar{u}^2 - 3 \left( \bar{u} - \frac{1}{3} \right)^2 \right]$ .

pour :  $2/3 \leq \bar{u} \leq 1$  ; la fonction de densité de  $\bar{u}$  est  $\frac{27}{2}(1-\bar{u})^2$ .

3.6. — La fonction de distribution de la moyenne logarithmique d'un échantillon de  $n$  variables indépendantes tirées d'une loi continue uniforme est une loi gamma incomplète dont le paramètre de position est nul, les paramètres d'échelle et de forme égaux à  $n$ , car  $\text{Log } u$  est distribué suivant une loi gamma incomplète de paramètres 0, 1, 1.

3.7. — Méthodes de Monte-Carlo.

3.7.1. — L'intérêt de la distribution continue uniforme vient de son utilisation dans les méthodes de Monte-Carlo, où l'on tire au hasard des valeurs comprises entre zéro et un, avec le nombre de chiffres significatifs choisis pour le problème à étudier. Chacune de ces valeurs est une valeur particulière de la variate continue uniforme, et est assimilable à une probabilité. De cette probabilité, on peut remonter, par inversion de la loi choisie pour le problème à étudier, à une valeur d'une variable aléatoire.

Par exemple, pour étudier correctement la distribution du coefficient d'aplatissement  $\gamma_2$  d'une loi de Gauss adaptée à un échantillon de taille  $n$ , il faudrait tirer  $n$  nombres au hasard compris entre zéro et un, inverser ces  $n$  probabilités d'une loi normale de paramètres de position = 0 et d'échelle = 1, d'où  $n$  variables aléatoires normales ; cet échantillon fournirait une valeur de  $\gamma_2$  (valeur indépendante des paramètres de position et d'échelle). En recommençant  $N$  fois l'opération totale, on aurait  $N$  valeurs de  $\gamma_2$  dont le simple classement permettrait l'établissement expérimental (avec une précision d'autant meilleure que  $N$  sera grand) de la distribution du coefficient d'aplatissement de la loi normale pour un échantillon de taille  $n$ .

3.7.2. — Génération de nombres au hasard.

La méthode la plus simple de génération de nombres « pseudo-aléatoires » est la méthode de congruence multiplicative :

$$x_i = a x_{i-1} \pmod{m},$$

$x_i$  est le reste de la division du produit ( $a$  multiplié par  $x_{i-1}$ ) par le diviseur  $m$ ,  $x_i$ ,  $x_{i-1}$ ,  $a$  et  $m$  étant des nombres entiers positifs.

La suite des nombres ainsi créés, compris entre 1 et  $(m - 1)$  est périodique, il est facile de tester qu'un échantillon tiré de cette suite, mais de taille bien inférieure à la période, peut être considéré comme formé de nombres au hasard.

I.B.M. propose le schéma suivant de génération de N nombres au hasard, compris entre 0 et 1, c'est-à-dire répartis suivant une distribution continue uniforme, valable pour les ordinateurs « 360 ».

```
Ligne 1      IX =
2           DO 3 I = 1,N
3           IX = IX* 65539
4           IF (IX) 1, 2, 2
5      1     IX = IX + 2147483647 + 1
6      2     PR (I) = DFLOAT (IX) / 2147483647.0
```

Le plus grand nombre entier qui puisse être retenu par les ordinateurs « 360 » est :  $2^{31} - 1 = 2147483647$  (nombre premier).

Si IX initial est supérieur ou égal à 32767, la ligne 3 donne le premier modulo demandé compris entre 3 et 2147483645, qui remplace la valeur précédente de IX, puis, divisé par  $2^{31} - 1$  à la ligne 6 donne un nombre compris entre 0 et 1.

Le multiplicateur 65539 est un nombre premier voisin de  $2^{16}$  et a été choisi après tests.

Pour que la période de la suite soit égale à  $2^{29}$  termes soit 536870912 termes, il suffit que le IX initial soit impair. Il est recommandé de le prendre avec 7 ou 8 chiffres.

## IV. — DISTRIBUTION DE GAUSS OU DISTRIBUTION NORMALE

4.1. — Nous ne nous étendrons pas sur la loi normale, la mieux connue et la plus étudiée des lois de probabilités usuelles.

4.1.1. — Fonction de répartition.

$$\text{Elle s'écrit : } F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du \quad \text{avec } u = \frac{x - x_0}{\sigma}.$$

$F(x)$  fréquence au non dépassement.

Lorsque  $u$  (ou  $x$ ) varie de  $-\infty$  à  $+\infty$ ,  $F(x)$  croît de zéro à un.  
 $x_0$  est le paramètre de position : valeur moyenne, ou mode, ou médiane.  
 $\sigma$  est le paramètre d'échelle : positif, différent de zéro, appelé écart-type.

4.1.2. — La fonction de densité est  $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$

dont la dérivée première s'écrit  $\frac{-1}{\sqrt{2\pi}} u e^{-\frac{u^2}{2}}$  racine  $u = 0$ ,

et la dérivée seconde  $\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} (u^2 - 1)$  racines  $u = \pm 1$ .

La fonction de densité est représentée par la courbe « en cloche » symétrique par rapport à  $u = 0$  qui est le mode, avec points d'inflexion à  $u = \pm 1$  (Figure 5).

4.1.3. — Moments, cumulants, valeurs centrales d'une distribution normale parfaitement connue.

Les moments d'ordre  $r$  de la variable réduite sont des moments centrés, ils n'existent que pour  $r \geq 1$ .

Les moments centrés d'ordre impair sont nuls.

Les moments centrés d'ordre «  $2r$  » pair sont  $\frac{1}{2^r} \frac{(2r)!}{r!}$ .

On en déduit les moments non centrés de la loi non réduite en faisant la transformation  $u = \frac{x - x_0}{\sigma}$  (paragraphe 1.3.2.).

Les cumulants de la loi réduite sont :

$$K_1 = 0, \quad K_2 = 1, \quad \text{et pour : } r \geq 3 \quad K_r = 0,$$

d'où les cumulants de la loi non réduite :

$$K_1 = x_0, \quad K_2 = \sigma^2, \quad \text{et pour : } r \geq 3 \quad K_r = 0.$$

La moyenne, le mode et la médiane ont pour valeur  $x_0$  dans la loi non réduite, la variance a pour valeur  $\sigma^2$ .

Les « erreurs-types » des valeurs centrales (racines carrées des variances respectives) pour un échantillon de taille  $n$  tiré d'une distribution normale dont les paramètres  $x_0$  et  $\sigma$  sont connus par avance, sont :

- erreur-type de la moyenne  $\sigma/\sqrt{n}$ ;
- erreur-type de la variance  $\sigma^2 \sqrt{\frac{2}{n}}$ ;

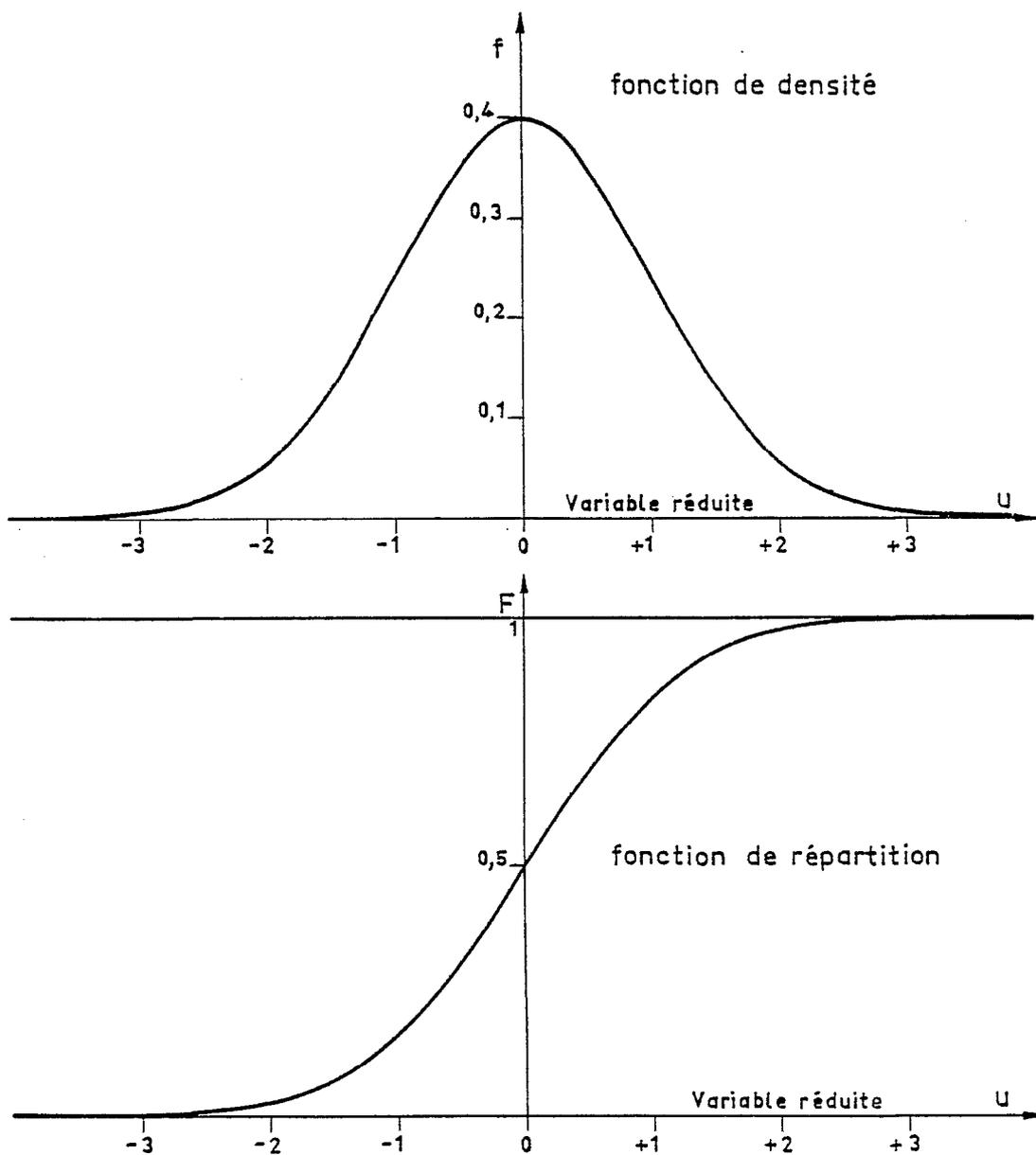


FIG. 5.

- erreur-type de la médiane  $1,2533 \cdot \sigma / \sqrt{n}$  ;
- erreur-type de l'écart-type  $\sigma / \sqrt{2n}$ .

Les moyennes harmoniques et géométriques n'ont pas de sens.

#### 4.2. — Distribution aléatoire de moments calculés.

La loi choisie étant une loi normale, dont les paramètres sont calculés d'après un échantillon de taille  $n$  (cf 4.3.).

##### 4.2.1. — Distribution de la moyenne.

$x_0$  étant la moyenne et  $s$  l'écart-type, calculés d'après l'échantillon, nous désirons connaître la probabilité qu'a une valeur  $x_m$  de représenter la moyenne de la population-mère.

La probabilité pour que la valeur absolue de  $t$ , défini par  $t = \frac{x_m - x_0}{s / \sqrt{v}}$  ( $v$  étant le nombre de degrés de liberté, ici  $v = n - 1$ ), soit inférieure à une valeur donnée, est une loi de Student, loi symétrique et moins aplatie que la loi normale, tendant assez rapidement vers une loi normale lorsque  $n$  augmente.

Cette probabilité, au non dépassement, est égale à la probabilité au dépassement, définie par une loi Bêta incomplète de paramètres  $p = \frac{v}{2}$ ,  $q = \frac{1}{2}$ , bornes zéro et un et de variate

$$u = \frac{v}{v + t^2}.$$

Notons que les tables que l'on trouve de la loi de Student donnent toujours la probabilité au dépassement.

##### 4.2.2. — Distribution de la variance.

$s^2$  étant la variance calculée d'après l'échantillon, nous désirons connaître la probabilité qu'a une valeur  $\sigma^2$  de représenter la variance de la population-mère (la moyenne n'intervient pas, bien que calculée d'après l'échantillon).

La probabilité pour que la valeur  $\chi^2 = \frac{\sigma^2}{s^2/n}$  soit inférieure à une valeur donnée est une loi de  $\chi^2$  à  $v$  (ici  $v = n - 1$ ) degrés de liberté, loi asymétrique tendant lentement vers une loi normale lorsque  $n$  augmente.

Cette probabilité, au non dépassement, est égale à la probabilité au non dépassement définie par une loi gamma incomplète de paramètres  $\gamma = \frac{v}{2}$ ,  $s = 1$ , de borne zéro et de variate  $u = \frac{\chi^2}{2}$ .

Notons que les tables que l'on trouve de la loi du  $\chi^2$  donnent toujours des probabilités au dépassement.

4.2.3. — Les distributions des moyennes et des variances dans des échantillons tirés d'une loi normale sont indépendantes : c'est une caractéristique de la loi normale, qui est la seule à posséder cette propriété.

Les distributions ci-dessus sont utilisées pour comparer des moyennes et des variances dans le cas d'échantillons provenant de lois supposées normales pour tester s'ils proviennent de la même population-mère. Les tables donnent directement le niveau de signification correspondant au risque de première espèce (cf paragraphe 2.3.3.).

#### 4.3. — Estimation des paramètres.

La loi choisie étant une loi normale, les paramètres sont estimés comme suit, d'après un échantillon de taille  $n$  :

Le paramètre de position est la moyenne :  $x_0 = \frac{1}{n} \sum_1^n x_i$ .

Le paramètre d'échelle  $\sigma$  est la racine carrée de la variance :

$$\sigma^2 = \frac{1}{n-1} \sum_1^n (x_i - x_0)^2.$$

Il est à noter que l'estimation de la variance par la méthode du maximum de vraisemblance conduit à :

$$\sigma^2 = \frac{1}{n} \sum_1^n (x_i - x_0)^2,$$

estimation « consistante » mais moins correcte que la précédente.

#### 4.4. — Test de normalité.

Il n'existe pas de test proprement dit de normalité. On peut se servir des coefficients d'asymétrie  $\gamma_1$  et d'aplatissement  $\gamma_2$  dont les distributions tendent vers des distributions normales si la taille  $n$  de l'échantillon est assez grande. Cette tendance est assez lente.

La valeur moyenne de  $\gamma_1$  est zéro et sa variance :

$$\frac{6n(n-1)}{(n-2)(n+1)(n+3)} \quad \text{tendant vers } \frac{6}{n}.$$

La valeur moyenne de  $\gamma_2$  est zéro et sa variance :

$$\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)} \quad \text{tendant vers } \frac{24}{n}$$

#### 4.5. — Comportement asymptotique de la distribution normale.

Si la fréquence  $F$  tend vers 1, et en posant  $T = \frac{1}{1-F}$ ,  $T$  étant le temps de récurrence, la variable réduite  $u$  tend vers  $\sqrt{2 [\text{Log } T - \text{Log } u - \text{Log } \sqrt{2\pi}]}$  c'est-à-dire que  $u$  tend vers  $\sqrt{2 \text{Log } T}$  d'une façon parabolique.

#### 4.6. — Calculs à l'ordinateur.

4.6.1. — Calcul de la probabilité correspondant à une valeur donnée de la variable.

L'ordinateur possède en bibliothèque la fonction  $\text{DERF}(Z) = \frac{2}{\sqrt{\pi}} \int_0^Z e^{-t^2} dt$  qui est disponible en double précision. Il suffit donc pour avoir la probabilité correspondant à une valeur  $x$ , les paramètres  $x_0$  et  $\sigma$  étant connus, de faire le changement de variable :

$$Z = \frac{x - x_0}{\sqrt{2} \sigma}$$

et la probabilité est :  $.5 + .5 * \text{DERF}(Z)$ .

4.6.2. — Calcul de la valeur de la variable correspondant à une probabilité donnée.

L'inversion de la distribution normale se fait assez facilement sur ordinateur en utilisant le procédé par corde et tangente et en se servant de la fonction de bibliothèque DERF (Z).

4.6.3. — Calcul des paramètres.

La sous-routine est facile à établir, et n'appelle aucun commentaire.

#### 4.7. — Anamorphose "normale".

4.7.1. — L'anamorphose normale revient à faire un changement de variable, transformant la variable d'origine en variable normale. On peut utiliser sur les échantillons de variables transformées les tests et les distributions de moments mis au point pour les distributions normales.

Le retour à la variable observée peut être difficile. Par exemple : il est facile de déterminer un intervalle de confiance de la moyenne des variables transformées, mais comment déterminerons-nous l'intervalle de confiance de la moyenne des variables originelles lorsqu'un paramètre de forme intervient dans leur distribution en plus des paramètres de position et d'échelle ?

4.7.2. — Principales transformations utilisées.

Distribution gaussio-logarithmique (cf. chapitre 6). Elle revient à faire le changement de variable :

$$\text{variable normale } y = \frac{1}{\sigma} \text{Log} \frac{x - x_0}{s} = \frac{y - y_0}{\sigma_0}$$

Distribution gamma incomplète (cf. chapitre 7). Si le paramètre de forme  $\gamma$  est assez grand et le paramètre de position nul, les racines cubiques (ou carrées si  $\gamma$  est très grand) d'une variate gamma incomplète suivent approximativement une loi normale non réduite.

Les transformations les plus générales sont :

- le développement de Cornish-Fischer qui fait intervenir les cumulants de la distribution originelle pour transformer la variable originelle en variable normale. Nous ne la citons que pour mémoire à cause de sa complication,
- la double anamorphose : on calcule la probabilité correspondant à chaque valeur de la variable observée dans la fonction de distribution adoptée pour l'échantillon observé, puis on inverse la fonction de distribution normale de façon à avoir la valeur de la variable normale réduite pour chaque probabilité calculée.

#### 4.8. — Approximations intéressantes.

Approximation de la probabilité pour une variable réduite  $u \geq 0$ .

$$F(u) \approx 1 - \frac{1}{2} \frac{1}{(1 + C_1 u + C_2 u^2 + C_3 u^3 + C_4 u^4)^4} \quad \text{avec une erreur} < 0,00025.$$

$$C_1 = 0,196854 \quad C_2 = 0,115194 \quad C_3 = 0,000344,$$

$$C_4 = 0,019527.$$

Approximation de la variable réduite pour une probabilité  $\geq 0,5$ .

$$u \approx t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \quad \text{où } t = \sqrt{\text{Log}(1/(1 - F(u))^2)}$$

$$C_0 = 2,515517, \quad C_1 = 0,802853, \quad C_2 = 0,010328,$$

$$d_1 = 1,432788, \quad d_2 = 0,189269, \quad d_3 = 0,001308.$$

L'erreur sur  $u$  reste inférieure à 0,00045.

## V. — DISTRIBUTION DE GUMBEL OU DOUBLEMENT EXPONENTIELLE

5.1. — Elle est dite aussi “loi des valeurs extrêmes”.

Considérons des échantillons de taille  $n$  de variables aléatoires indépendantes, et une valeur  $u$  de la variate dont la probabilité au non dépassement est  $F(u)$ . La probabilité pour que toutes les valeurs d'un échantillon soient inférieures à  $u$  est  $[1 - (1 - F(u))^n]$ .

Si  $n$  est grand et  $1 - F(u)$  petit, la probabilité ci-dessus peut s'écrire approximativement, exp. :  $[-n(1 - F(u))]$ .

En supposant que  $n[1 - F(u)]$  puisse se mettre sous la forme  $e^{-u}$ , ce qui est purement gratuit, la fonction de probabilité devient  $e^{-e^{-u}}$ .

5.2. — Expressions mathématiques.

5.2.1. — Fonction de répartition.

Nous l'écrivons  $F(x) = \frac{1-S}{2} + S e^{-e^{-u}}$  avec  $u = \frac{x-x_0}{s}$  et  $S = 1$  avec le signe de  $s$ .

Soit  $F(x)$  la fréquence au non dépassement :

Lorsque  $x$  varie de  $-\infty$  à  $+\infty$ ,  $F(x)$  croit de zéro à un.

$x_0$  est le paramètre de position : mode.

$s$  est le paramètre d'échelle, différent de zéro, positif (ou négatif).

5.2.2. — La fonction de densité est  $f(u) = e^{-u} e^{-e^{-u}}$  dont la dérivée première s'écrit  $e^{-u} e^{-e^{-u}} (e^{-u} - 1)$ , de racine  $u = 0$ , et la dérivée seconde  $e^{-u} e^{-e^{-u}} (1 - 3e^{-u} + e^{-2u})$ , de racines  $u = \text{Log} \left[ \frac{1}{2} (3 \pm \sqrt{5}) \right]$ .

La fonction de densité est représentée par une courbe en cloche asymétrique (asymétrie du signe de  $s$ ), avec points d'inflexion à  $u = \pm 0,96243$  (Figure 6).

5.2.3. — Moments, cumulants, valeurs centrales d'une distribution de Gumbel parfaitement connue.

Les cumulants d'ordre  $\geq 2$  sont, pour la variable réduite :

$$K_r = (r-1)! \zeta(r), \quad \text{fonction de Rieman } \zeta(r) = \sum_{i=1}^{\infty} \frac{1}{i^r}$$

comme :

$$\begin{array}{ll} \zeta(2) = \frac{\pi^2}{6} & \text{il vient : } K_2 = 1,644934, \\ \zeta(3) = 1,2020569 & K_3 = 2,404114, \\ \zeta(4) = 1,0823232 & K_4 = 6,493939. \end{array}$$

Les cumulants sont, pour la variable non réduite :

$$\begin{array}{l} K_1 = x_0 + s \times 0,5772157 \\ K_2 = s^2 \times 1,644934 \\ K_3 = s^3 \times 2,404114 \\ K_4 = s^4 \times 6,493939 \end{array}$$

De ces cumulants on peut tirer, si besoin, les moments centrés ou non (cf. paragraphe 1.3.3.):

La moyenne a pour valeur  $x_0 + s \times 0,5772157$ .

La probabilité correspondante est 0,570, si  $s > 0$ , et 0,430, si  $s < 0$ .

Le mode est  $x_0$ , la probabilité correspondante est 0,368, si  $s > 0$ , et 0,632, si  $s < 0$ .

La médiane a pour valeur  $x_0 + s \times 0,3665131$ .

La variance a pour valeur  $s^2 \times 1,644934$ .

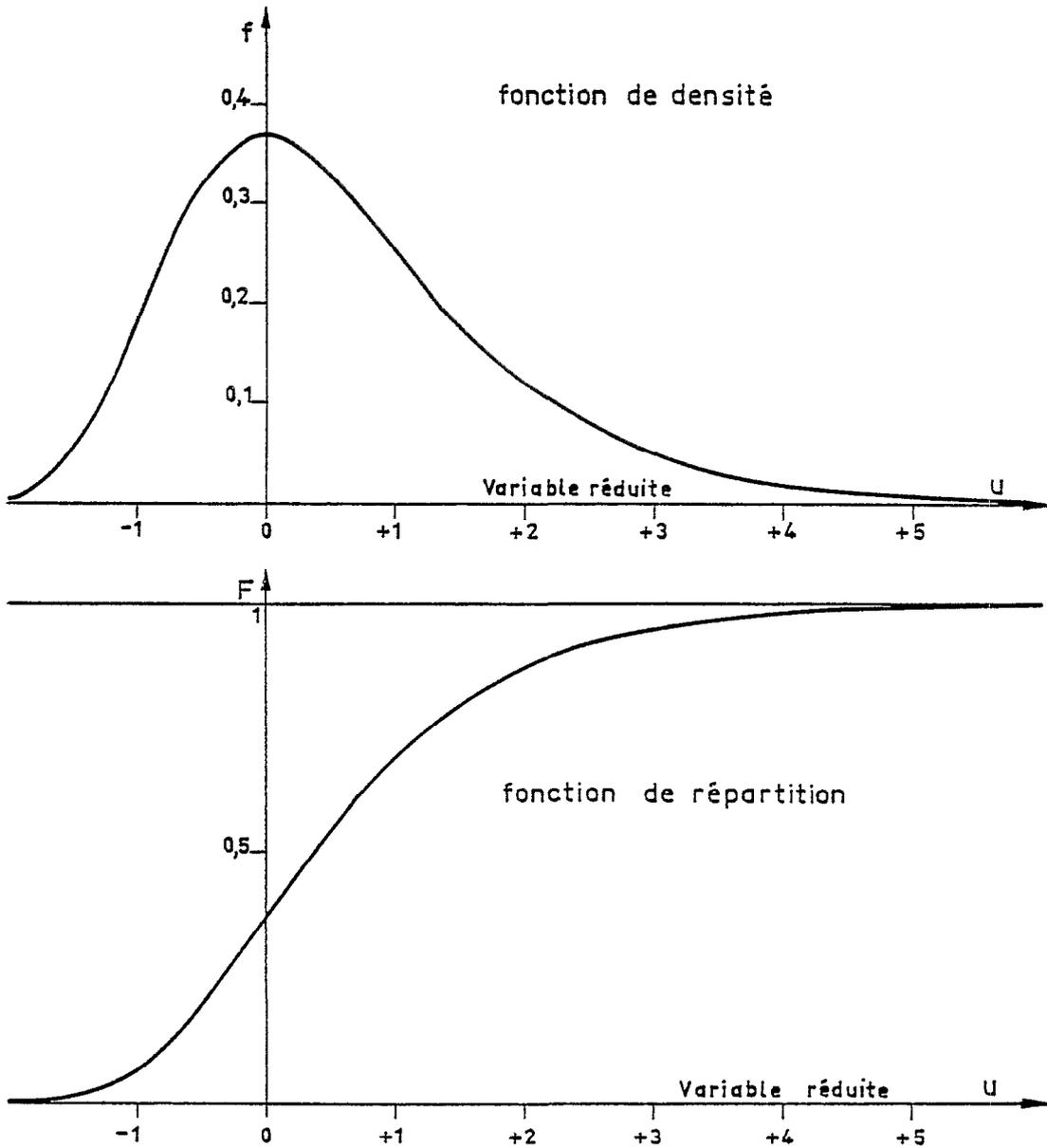


FIG. 6.

Le coefficient d'asymétrie vaut 1,139 avec le signe de  $s$ ; le coefficient d'aplatissement a pour valeur 2,400 : la distribution est moins aplatie que la normale.

Les moyennes harmonique et géométrique n'ont pas de sens.

### 5.3. — Estimation des paramètres par les moments (échantillon de taille $n$ ).

En utilisant les sommes des trois premières puissances de  $x$  et en posant :

$$S_1 = \frac{n}{1} \sum x_i, \quad S_2 = \frac{n}{1} \sum x_i^2, \quad \text{et} \quad S_3 = \frac{n}{1} \sum x_i^3,$$

il vient :

$$s^2 = (0,780)^2 \times \frac{1}{n-1} \left( S_2 - \frac{1}{n} S_1^2 \right).$$

et :

$$x_0 = \frac{1}{n} S_1 - s \times 0,5772.$$

le signe de  $s$  étant celui de (moyenne — médiane) ou, avec une meilleure certitude si la taille de l'échantillon n'est pas grande, déterminé par le signe de :  $K_3 = S_3 - \frac{3}{n} S_2 S_1 + \frac{2}{n^2} S_1^3$ .

### 5.4. — Détermination des paramètres par la méthode du maximum de vraisemblance (échantillon de taille $n$ ).

Pour une valeur  $x_i$  de la variable, la fonction de densité a comme valeur :

$$\frac{1}{|s|} e^{-\frac{x_i - x_0}{s}} e^{-e^{-\frac{x_i - x_0}{s}}}$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$-n \operatorname{Log} |s| - \sum \frac{x_i - x_0}{s} - \sum e^{-\frac{x_i - x_0}{s}}.$$

en dérivant par rapport à  $x_0$  et à  $s$ , et en annulant les dérivées :

$$s + \sum x_i e^{-\frac{x_i}{s}} / \sum e^{-\frac{x_i}{s}} = \frac{1}{n} \sum x_i,$$

$$\frac{x_0}{s} = \operatorname{Log} n - \operatorname{Log} \left( \sum e^{-\frac{x_i}{s}} \right).$$

Si nous considérons :  $Y = s + \sum x_i e^{-\frac{x_i}{s}} / \sum e^{-\frac{x_i}{s}}$ , indépendant de  $x_0$ , nous constatons que la dérivée de  $Y$  par rapport à  $s$  est toujours supérieure à 1 ; la liaison entre  $Y$  et  $s$  est donc monotone. La sous-routine de calcul des paramètres par un ordinateur est assez facile à écrire, mais il faut commencer par déterminer le signe de  $s$  par  $K_3$  comme au paragraphe 5.3.

### 5.5. — Comportement asymptotique de la distribution de Gumbel.

Si la fréquence  $F$  tend vers 1, et en posant :  $T = \frac{1}{1-F}$ ,  $T$  étant le temps de récurrence ou période de retour, la variable réduite  $u$  tend vers  $\operatorname{Log} T$ , si  $s > 0$ , et  $|u|$  tend vers  $\operatorname{Log}(\operatorname{Log} T)$ , si  $s < 0$ .

### 5.6. — Calculs à l'ordinateur.

Les paramètres  $x_0$  et  $s$  étant connus, il est facile d'établir les sous-programmes de calcul d'une probabilité correspondant à une valeur donnée de la variable, en utilisant la fonction de bibliothèque  $\operatorname{EXP}(Z)$ , ou d'inversion de la fonction en utilisant la fonction de bibliothèque  $\operatorname{ALOG}(Z)$ .

## VI. — DISTRIBUTION GAUSSO-LOGARITHMIQUE

### 6.1. — Expressions mathématiques.

#### 6.1.1. — Fonction de répartition.

Nous l'écrivons : 
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^u \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} du$$

avec : 
$$u = \frac{x - x_0}{s},$$

formulation qui fait bien ressortir les rôles des paramètres, plutôt que d'utiliser l'écriture :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u'} e^{-\frac{u'}{2}} du', \quad \text{avec : } u' = a \text{Log}(x - x_0) + b$$

$F(x)$  fréquence au non dépassement, lorsque  $u$  varie de 0 à  $+\infty$  (ou  $x$  de  $x_0$  à  $+\infty$ ),  $F(x)$  croît de zéro à un.

$x_0$  est le paramètre de position : borne inférieure.

$s$  est le paramètre d'échelle, positif, différent de zéro.

$\sigma$  est le paramètre de forme, positif, différent de zéro.

6.1.2. — La fonction de densité est :  $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2\sigma^2}}$ , dont la dérivée première s'écrit :  $\frac{-1}{\sigma\sqrt{2\pi}} \frac{1}{u^2} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} \left(1 + \frac{\text{Log} u}{\sigma^2}\right)$ , de racine  $\text{Log} u = -\sigma^2$  et la dérivée seconde :  $\frac{1}{\sigma\sqrt{2\pi}} \frac{1}{u^3} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} \left(\frac{\text{Log}^2 u}{\sigma^4} + \frac{3 \text{Log} u}{\sigma^2} + 2 - \frac{1}{\sigma^2}\right)$ , de racines  $\text{Log} u = \frac{-\sigma^2}{2} \left(3 \pm \sqrt{1 + \frac{4}{\sigma^2}}\right)$ .

La fonction de densité est représentée par une courbe en cloche asymétrique. La tangente en  $u = 0$  est toujours nulle, et il y a toujours deux points d'inflexion. Quel que soit  $u$ , les valeurs  $u$  et  $\frac{e^{-2\sigma^2}}{u}$ , ont la même densité de probabilité (figure 7).

6.1.3. — Moments, cumulants, valeurs centrales d'une distribution gaussio-logarithmique parfaitement connue.

Le moment d'ordre  $r$  de la variable réduite est  $m_r = e^{\frac{r^2 \sigma^2}{2}}$ , quel que soit  $r$  positif ou négatif dont on peut déduire le moment (non centré) d'ordre  $r$  de la variable non réduite en faisant la transformation  $u = \frac{x - x_0}{s}$  (cf paragraphe 1.3.2.).

La moyenne est :  $x_0 + s e^{\frac{\sigma^2}{2}}$  ;

La variance est :  $K_2 = s^2 e^{\sigma^2} (e^{\sigma^2} - 1)$  ;

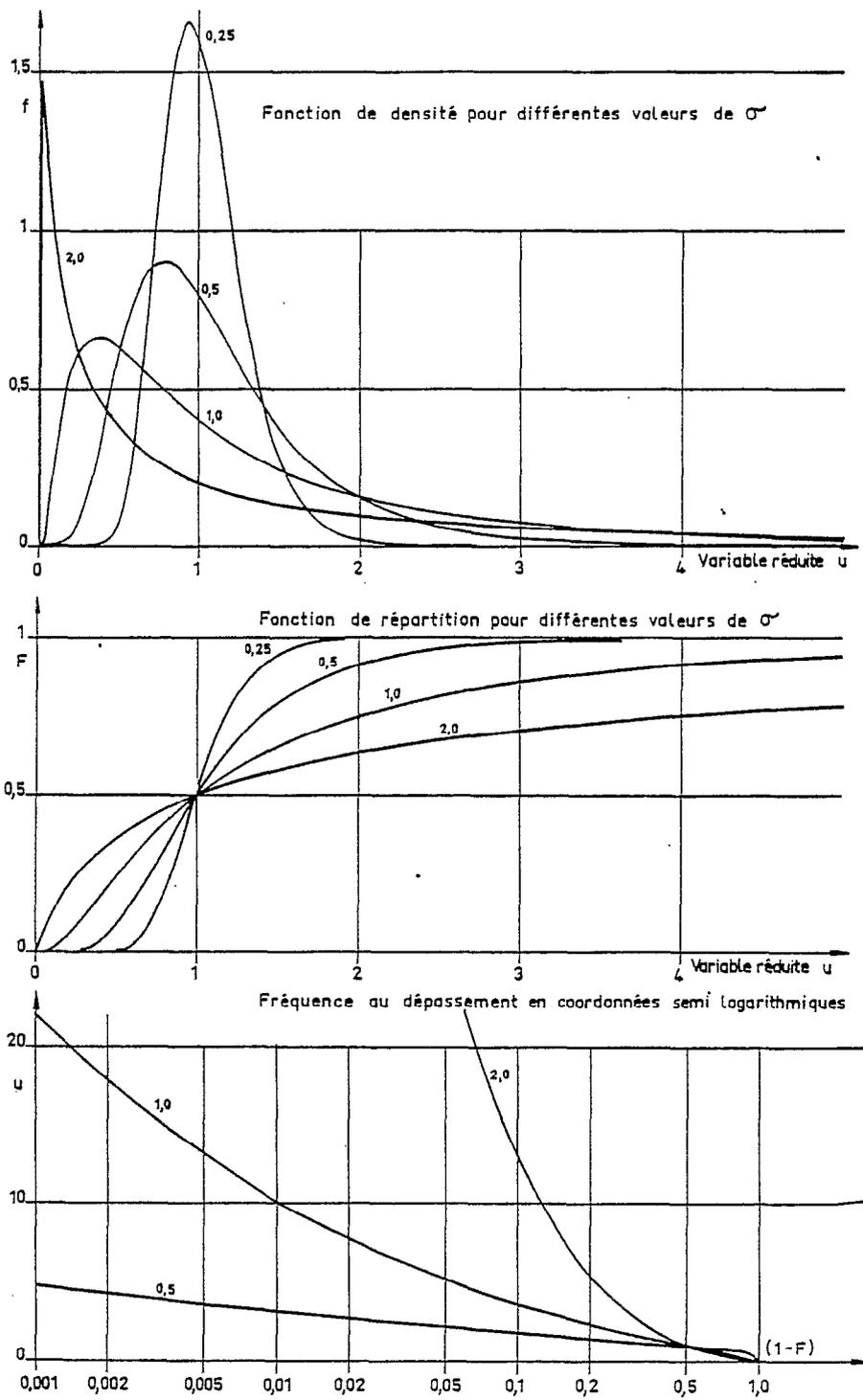


FIG. 7.

Le cumulatif d'ordre 3 est :  $K_3 = s^3 e^{3\sigma^2/2} (e^{\sigma^2} - 1)^2 (e^{\sigma^2} + 2)$  ;

Le cumulatif d'ordre 4 est :  $K_4 = s^4 e^{2\sigma^2} (e^{\sigma^2} - 1)^3 (e^{3\sigma^2} + 3 e^{2\sigma^2} + 6 e^{\sigma^2} + 6)$ .

Le coefficient d'asymétrie est :  $\gamma_1 = (e^{\sigma^2} + 2) \sqrt{e^{\sigma^2} - 1}$ ,

L'asymétrie est toujours positive et augmente avec  $\sigma$  ;

Le coefficient d'aplatissement est :  $\gamma_2 = (e^{\sigma^2} - 1) (e^{3\sigma^2} + 3 e^{2\sigma^2} + 6 e^{\sigma^2} + 6)$ . Il est toujours positif, et la distribution est d'autant moins aplatie par rapport à la normale que  $\sigma$  est plus grand. Lorsque  $\sigma$  tend vers zéro, la distribution tend vers une distribution normale, les deux coefficients  $\gamma_1$  et  $\gamma_2$  tendant vers zéro (Figure 8).

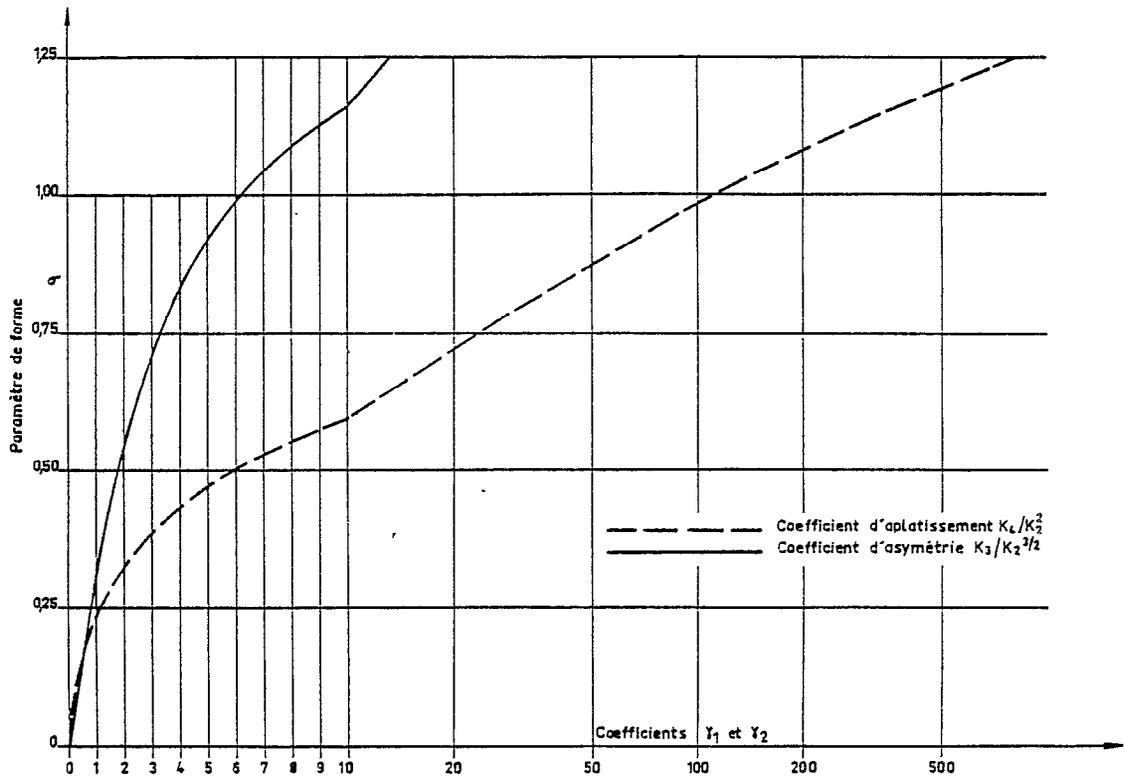


FIG. 8.

Le mode est :  $x_0 + s e^{-\sigma^2}$  ;

La médiane est :  $x_0 + s$ .

La moyenne harmonique de la variable réduite est  $e^{-\frac{\sigma^2}{2}}$ , l'inverse de la moyenne harmonique est le moment d'ordre  $-1$  dont la variance est la même que celle de la moyenne.

La moyenne logarithmique de la variable réduite est nulle, donc la moyenne géométrique est égale à 1 quel que soit  $\sigma$ .

Le graphique de la figure 9 montre les positions relatives, suivant les valeurs de  $\sigma$  et pour des lois gaussio-logarithmiques réduites, du mode, de la médiane, et des moyennes arithmétique, géométrique et harmonique.

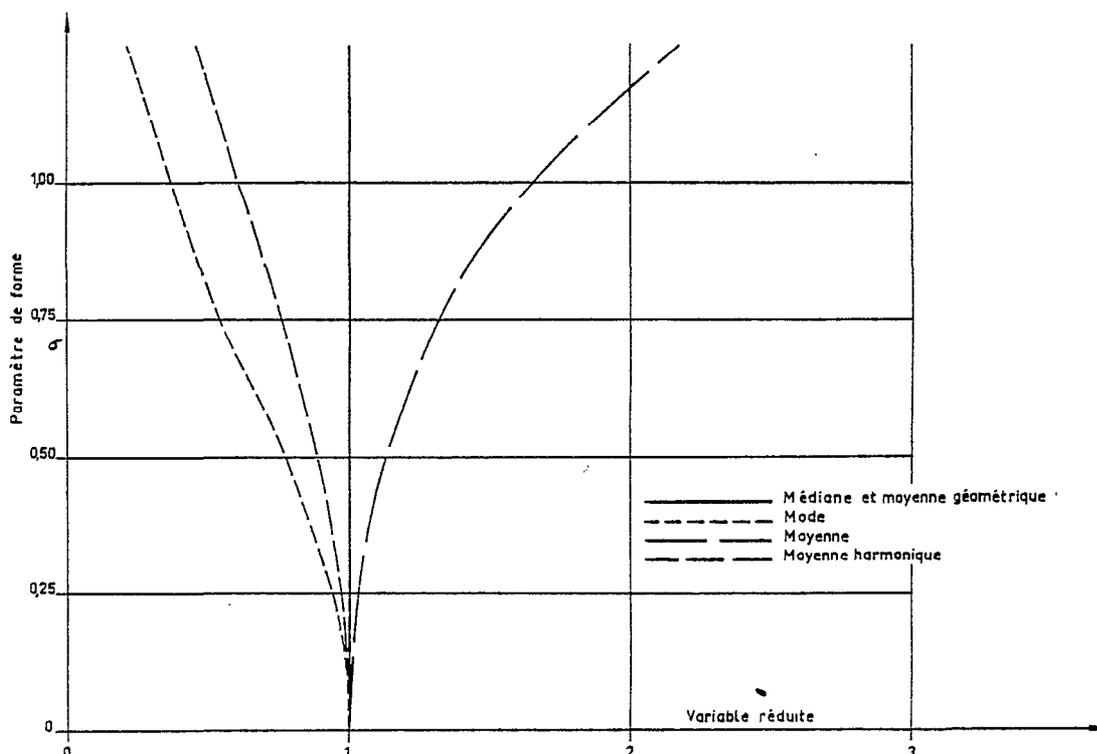


FIG. 9.

6.2. — Estimation des paramètres par les moments (échantillon de taille  $n$ ).

En utilisant les sommes des trois premières puissances des valeurs observées ( $x_1 \dots x_i \dots x_n$ ) et en posant :

$$S_1 = \frac{1}{n} \sum x_i, \quad S_2 = \frac{1}{n} \sum x_i^2, \quad S_3 = \frac{1}{n} \sum x_i^3, \quad A = e^{\frac{\sigma^2}{2}}$$

on obtient les équations :

$$K_1 = s A = S_1 - x_0,$$

$$K_2 = s^2 A^2 (A^2 - 1) = \frac{n}{n-1} (S_2 - S_1^2),$$

$$K_3 = s^3 A^3 (A^2 - 1)^2 (A^2 + 2) = \frac{n^2}{(n-1)(n-2)} (S_3 - 3 S_2 S_1 + 2 S_1^3).$$

Nous proposons la solution suivante :

$$K_3^2 / K_2^3 = (A^2 - 1) (A^2 + 2)^2 = \frac{n(n-1)}{(n-2)^2} \times \frac{(S_3 - 3 S_2 S_1 + 2 S_1^3)^2}{(S_2 - S_1^2)^3} = 2 C \quad \text{avec } C > 0.$$

L'équation du 3<sup>e</sup> degré en  $A^2$  a une racine réelle unique et positive qui est :

$$A^2 = (1 + C + \sqrt{2 C + C^2})^{1/3} + (1 + C - \sqrt{2 C + C^2})^{1/3} - 1,$$

d'où  $\sigma$  par  $\sigma^2 = 2 \text{ Log } A$  ;

$$s \text{ par } s^2 = \frac{S_2 - S_1^2}{A^2(A^2 - 1)} \times \frac{n}{n-1};$$

$$x_0 \text{ par } x_0 = S_1 - s A.$$

6.2.1. — Si  $x_0$  est choisi *a priori*, il est plus simple de passer par les rapports des moments non centrés :

$$R_1 = \frac{\Sigma (x_i - x_0)}{n} = s e^{\frac{1}{2} \sigma^2},$$

$$R_2 = \frac{\Sigma (x_i - x_0)^2}{\Sigma (x_i - x_0)} = s e^{\frac{3}{2} \sigma^2}.$$

$$R_3 = \frac{\Sigma (x_i - x_0)^3}{\Sigma (x_i - x_0)^2} = s e^{\frac{5}{2} \sigma^2},$$

d'où :

$$e^{\sigma^2} = \frac{R_2}{R_1} \quad \text{et} \quad \sigma^2 = \text{Log } R_2 - \text{Log } R_1,$$

et :

$$s = R_1 e^{-\frac{1}{2} \sigma^2} = R_1 \sqrt{\frac{R_1}{R_2}}.$$

Remarquons que la quantité  $R = \text{Log } R_1 + \text{Log } R_3 - 2 \text{Log } R_2$ , qui n'est pas forcément nulle pour un échantillon de taille finie, a une valeur moyenne nulle pour un grand nombre d'échantillons de la même taille tirés de la même population-mère. En supposant linéaire la corrélation qui existe entre la quantité  $R$  et la quantité  $R' = \text{Log } R_2 - \text{Log } R_1$  nous pourrions améliorer la détermination de  $\sigma^2$  d'une façon analogue à la méthode suivie dans la loi gamma incomplète (cf. paragraphe 7.2.2.). Les expressions auxquelles on aboutit sont très compliquées.

### 6.3. — Détermination des paramètres par le maximum de vraisemblance (échantillon de taille $n$ ).

Pour une valeur  $x_i$  de la variable la fonction de densité s'écrit :

$$\frac{1}{\sigma \sqrt{2\pi}} \frac{1}{x_i - x_0} e^{-\frac{\text{Log}^2 \left( \frac{x_i - x_0}{s} \right)}{2 \sigma^2}}.$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$-n \text{Log} (\sqrt{2\pi}) - n \text{Log } \sigma - \Sigma \text{Log} (x_i - x_0) - \frac{1}{2 \sigma^2} [\Sigma \text{Log}^2 (x_i - x_0) - 2 \text{Log } s \Sigma \text{Log} (x_i - x_0) + n \text{Log}^2 s].$$

En dérivant par rapport à  $x_0$ ,  $s$  et  $\sigma$  et en annulant les dérivées, on obtient après simplifications :

$$n - \frac{1}{\sigma^2} [\Sigma \text{Log}^2 (x_i - x_0) - 2 \text{Log } s \Sigma \text{Log} (x_i - x_0) + n \text{Log}^2 s] = 0;$$

$$n \text{Log } s - \Sigma \text{Log} (x_i - x_0) = 0;$$

$$\Sigma \frac{1}{x_i - x_0} - \frac{1}{\sigma^2} \left[ \text{Log } s - \Sigma \frac{1}{x_i - x_0} - \Sigma \frac{\text{Log} (x_i - x_0)}{x_i - x_0} \right] = 0;$$

en posant :

$$SH = \frac{1}{n} \Sigma \frac{1}{x_i - x_0},$$

$$SL = \frac{1}{n} \Sigma \text{Log} (x_i - x_0),$$

$$SLH = \frac{1}{n} \Sigma \frac{\text{Log} (x_i - x_0)}{x_i - x_0},$$

$$SLL = \frac{1}{n} \Sigma \text{Log}^2 (x_i - x_0),$$

il vient :

$$\text{Log } s = SL \quad \text{donnant } s \text{ une fois } x_0 \text{ déterminé,}$$

$$\sigma^2 = SLL - SL^2 \text{ donnant } \sigma \text{ une fois } x_0 \text{ déterminé,}$$

et  $SL - \frac{SLH}{SH} - SLL + SL^2 = 0$  où la seule inconnue est  $x_0$ .

Pour résoudre cette dernière équation, transcendante et implicite, nous posons :

$$SQ = SLL - SL^2,$$

$$SP = SL - SLH/SH.$$

La dérivée de SP par rapport à  $x_0$  est supérieure à la dérivée de SQ par rapport à  $x_0$ . Pour résoudre l'équation en  $x_0$ , il suffit de choisir une valeur initiale de  $x_0$ , calculer SP — SQ. Si cette quantité est négative, la valeur choisie pour  $x_0$  est trop petite, sinon la valeur choisie pour  $x_0$  est trop grande. D'où l'écriture de la sous-routine de détermination des paramètres à l'aide d'un ordinateur.

#### 6.4. — Estimation des paramètres : loi gaussio-logarithmique tronquée avec troncature. Estimation par les moments.

Dans le cas de la loi gaussio-logarithmique, nous ne trouvons jamais de forme en « J » (cf. paragraphe 7.4.). Mais lorsque le paramètre de forme devient supérieur à 1, on peut se trouver, dans la pratique, dans un cas analogue : le mode étant voisin de la borne inférieure  $x_0$ , le nombre des observations, qui auraient dû être faites dans la tranche  $x_0, x_h$  ( $x_h$  seuil de troncature, qui peut être supérieur au mode) et dont la fréquence totale n'est pas négligeable, est inconnu ainsi que les valeurs individuelles de ces observations.

Nous pouvons admettre (cf. paragraphe 7.4.) que la borne inférieure  $x_0$  est connue par avance avec une approximation suffisante. En posant :

$$S_1 = \Sigma (x_i - x_0), \quad S_2 = \Sigma (x_i - x_0)^2, \quad S_3 = \Sigma (x_i - x_0)^3, \quad S_4 = \Sigma (x_i - x_0)^4.$$

sommes qui sont convenablement connues car les observations non faites ( $x_i$  peu différent de  $x_0$ ) et donc non utilisées interviennent pour peu de chose dans ces totaux. En effet, nous nous servons de toutes les observations disponibles sans faire de troncature à un seuil supérieur à celui qui est imposé par la sensibilité de l'appareil de mesure.

Nous écrivons les équations :

$$R_2 = S_2/S_1 = s e^{\frac{3}{2}\sigma^2},$$

$$R_3 = S_3/S_2 = s e^{\frac{5}{2}\sigma^2},$$

$$R_4 = S_4/S_3 = s e^{\frac{7}{2}\sigma^2}.$$

Il suffit en principe des deux premières pour calculer les paramètres  $\sigma$  et  $s$  :

$$\sigma^2 = \text{Log } R_3 - \text{Log } R_2,$$

$$s = R_2 e^{-\frac{3}{2}\sigma^2} = \frac{R_2^{5/2}}{R_3^{3/2}}.$$

Le nombre « théorique » d'observations est M donné par :

$$M = S_1 / \left( s e^{\frac{\sigma^2}{2}} \right) = S_3 \left( \frac{S_1}{S_2} \right)^3$$

et si l'on appelle A le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, le paramètre  $F_0$  est calculé par :

$$F_0 = \frac{A - M}{A}.$$

On pourrait définir comme plus haut une quantité :

$$R = \text{Log } R_4 + \text{Log } R_2 - 2 \text{Log } R_3,$$

de valeur moyenne nulle, et utiliser la corrélation, supposée linéaire entre la quantité R et la quantité  $R' = \text{Log } R_3 - \text{Log } R_2$  pour améliorer la détermination de  $\sigma^2$ . Les expressions auxquelles on aboutit sont très compliquées.

#### 6.5. — Détermination des paramètres par le maximum de vraisemblance. Loi gaussio-logarithmique tronquée avec troncature.

La borne inférieure  $x_0$  étant connue par avance et le seuil de troncature  $x_h$  étant choisi (cf. paragraphe précédent).

On applique la méthode aux  $n$  observations conservées ( $x_i \geq x_h$ ) dont la probabilité totale est :

$$\frac{1 - F_0}{\sigma \sqrt{2} \pi} \int_{x_h}^{\infty} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2 \sigma^2}} du \quad \text{avec } u = (x - x_0)/s.$$

Pour une valeur  $x_i$  de la variable, la fonction de densité s'écrit :

$$\frac{1 - F_0}{\sigma \sqrt{2} \pi} \frac{1}{x_i - x_0} e^{-\frac{\text{Log}^2((x_i - x_0)/s)}{2 \sigma^2}} \bigg/ \frac{1 - F_0}{\sigma \sqrt{2} \pi} \int_{x_h}^{\infty} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2 \sigma^2}} du.$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$\begin{aligned} & -n \text{Log } \sigma - n \text{Log } \sqrt{2} \pi - \Sigma \text{Log } (x_i - x_0) \\ & \quad - \frac{1}{2 \sigma^2} [\Sigma \text{Log}^2 (x_i - x_0) - 2 \text{Log } s \Sigma \text{Log } (x_i - x_0) + n \text{Log}^2 s] \\ & \quad - n \text{Log} \left[ 1 - \frac{1}{\sigma \sqrt{2} \pi} \int_{x_0}^{x_h} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2 \sigma^2}} du \right]. \end{aligned}$$

En dérivant par rapport à  $s$  et à  $\sigma$  et en annulant les dérivées on obtient (dérivée par rapport à  $s$ ) :

$$\begin{aligned} & \frac{1}{s \sigma^2} \Sigma \text{Log } (x_i - x_0) - \frac{n}{s \sigma^2} \text{Log } s - n \left[ \frac{1}{s \sigma} \int_{x_0}^{x_h} \frac{-1}{u \sqrt{2} \pi} \frac{\text{Log } u}{\sigma^2} e^{-\frac{\text{Log}^2 u}{2 \sigma^2}} du \right] \bigg/ \\ & \quad \left[ 1 - \frac{1}{\sigma \sqrt{2} \pi} \int_{x_0}^{x_h} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2 \sigma^2}} du \right] = 0. \end{aligned}$$

qui s'écrit en posant :  $\frac{1}{n} \Sigma \text{Log } (x_i - x_0) = \text{SL}$ ,  $\frac{1}{n} \Sigma \text{Log}^2 (x_i - x_0) = \text{SM}$ ,

$$\text{et : } \frac{1}{\sigma \sqrt{2\pi}} \int_{x_0}^{x_h} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} du = P_h, \quad \frac{x_h - x_0}{s} = u_h,$$

$$\frac{1}{\sigma} (\text{SL} - \text{Log } s) = \frac{e^{-\frac{\text{Log}^2 u_h}{2\sigma^2}}}{\sqrt{2\pi} (1 - P_h)},$$

et (dérivée par rapport à  $\sigma$ ) :

$$\frac{-n}{\sigma} + \frac{1}{\sigma^3} \left[ \Sigma \text{Log}^2 (x_i - x_0) - 2 \text{Log } s \Sigma \text{Log} (x_i - x_0) + n \text{Log}^2 s \right]$$

$$- n \left[ \frac{1}{\sigma} \int_{x_0}^{x_h} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} \frac{du}{u} - \frac{1}{\sigma \sqrt{2\pi}} \int_{x_0}^{x_h} \frac{\text{Log}^2 u}{\sigma^3} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} \frac{du}{u} \right] / (1 - P_h) = 0,$$

qui peut s'écrire :

$$\frac{1}{\sigma^2} (\text{SM} - 2 \text{SL} \text{Log } s + \text{Log}^2 s) = 1 + \frac{\text{Log } u_h}{\sigma} \frac{e^{-\frac{\text{Log}^2 u_h}{2\sigma^2}}}{\sqrt{2\pi} (1 - P_h)}.$$

Nous nous trouvons en présence d'un système de deux équations implicites et transcendantes en  $s$  et  $\sigma$ , que l'on simplifie en gardant la première et une combinaison des deux :

$$\text{SM} - \text{SL} \text{Log} (x_h - x_0) - \sigma^2 = \text{Log } s [\text{SL} - \text{Log} (x_h - x_0)].$$

La méthode que nous pensons la meilleure pour le calcul par ordinateur consiste à choisir au départ une valeur de  $\sigma$ , d'où une valeur de  $s$  d'après la dernière équation, et faire varier  $\sigma$  pour que  $\text{SL}$  soit peu différent de :  $\text{Log } s + \sigma e^{-\frac{\text{Log}^2 u_h}{2\sigma^2}} / \sqrt{2\pi} (1 - P_h)$ .

Le paramètre de tronquage  $F_0$  est calculé par ( $A$  étant le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, et  $P_h$  étant la dernière valeur calculée plus haut de :

$$\frac{1}{\sigma \sqrt{2\pi}} \int_{x_0}^{x_h} \frac{1}{u} e^{-\frac{\text{Log}^2 u}{2\sigma^2}} du,$$

$n$  étant le nombre d'observations conservées, de valeurs égales ou supérieures au seuil de troncature) :

$$F_0 = 1 - \frac{n/A}{1 - P_h}.$$

## 6.6. — Comportement asymptotique de la distribution gaussio-logarithmique.

Si la fréquence  $F$  tend vers 1, et en posant  $T = \frac{1}{1 - F}$ ,  $T$  étant le temps de récurrence, la variable réduite  $u$  tend vers :  $\exp \sqrt{2} [\text{Log } T - \text{Log} (\text{Log } u) + \text{Log} (1/\sqrt{2\pi})] \sigma^2$ . Il n'y a pas d'asymptote à proprement parler, mais on peut dire que  $u$  tend vers  $e^{\sigma \sqrt{2 \text{Log } T}}$  d'une façon parabolique.

## 6.7. — Calculs à l'ordinateur.

Pour le calcul de la probabilité correspondant à une valeur donnée de la variable, connaissant les paramètres  $x_0$ ,  $s$  et  $\sigma$ , il suffit de faire la transformation :

$$y = \frac{\text{Log} ((x_i - x_0)/s)}{\sigma} \text{ pour être ramené au cas de la variable normale réduite : cf paragraphe 4.6.}$$

L'inversion de la loi gaussio-logarithmique se fait suivant la même marche que l'inversion de la loi normale, avec *in fine* la transformation de la variable normale en variable gaussio-logarithmique.

## VII. — DISTRIBUTION GAMMA INCOMPLÈTE

### 7.1. — Expressions mathématiques.

#### 7.1.1. — Fonction de répartition :

Elle s'écrit :

$$F(x) = \frac{1}{\Gamma(\gamma)} \int_0^u u^{\gamma-1} e^{-u} du \quad \text{avec } u = (x - x_0)/s$$

$F(x)$  est la fréquence au non-dépassement.

Lorsque  $u$  varie de 0 à  $+\infty$  (ou  $x$  de  $x_0$  à  $+\infty$ ),  $F(x)$  croît de zéro à un.

$\Gamma(\gamma)$  est la fonction gamma complète :  $\Gamma(\gamma) = \int_0^{\infty} u^{\gamma-1} e^{-u} du$  ;

$x_0$  est le paramètre de position, borne inférieure de l'intervalle de définition de la variate ;

$s$  est le paramètre d'échelle ;  $s$  ayant les mêmes dimensions que  $x$  et  $x_0$ , nous ne conserverons pas la notation  $a = 1/s$  prise d'habitude pour ce paramètre.  $s$  est positif, différent de zéro ;

$\gamma$  est le paramètre de forme, positif, différent de zéro.

7.1.2. — La fonction de densité est :  $f(u) = \frac{1}{\Gamma(\gamma)} u^{\gamma-1} e^{-u}$

dont la dérivée première s'écrit :  $\frac{1}{\Gamma(\gamma)} u^{\gamma-2} e^{-u} [(\gamma-1) - u]$ , de racine  $u = \gamma - 1$

et la dérivée seconde :  $\frac{1}{\Gamma(\gamma)} u^{\gamma-3} e^{-u} [(\gamma-2)(\gamma-1) - 2(\gamma-1)u + u^2]$ ,

de racines  $u = (\gamma-1) \pm \sqrt{\gamma-1}$ , d'où les différentes formes possibles de la représentation graphique de la fonction de densité.

$0 < \gamma < 1$  densité de probabilité infinie à  $u=0$ , pas de mode réel ni de points d'inflexion (mode observé à  $u=0$ ). La densité de probabilité tend vers zéro si  $u$  tend vers  $+\infty$ . Loi dite en « J » ;

$\gamma = 1$  densité de probabilité 1 à  $u=0$ , pas de mode réel ni de point d'inflexion (mode observé à  $u=0$ ). Loi exponentielle ;

$1 < \gamma < 2$  densité de probabilité nulle à  $u=0$ , avec tangente  $+\infty$ . Mode à  $u = \gamma - 1$ . Un point d'inflexion à  $u = \gamma - 1 + \sqrt{\gamma - 1}$  ; la seconde racine est négative en dehors de l'intervalle de définition de la variable ;

$\gamma = 2$  densité de probabilité nulle à  $u=0$  avec tangente  $+1$ . Mode à  $u = 1$ . Un point d'inflexion à  $u = 2$  (l'autre étant à  $u = 0$ ) ;

$2 < \gamma$  densité de probabilité nulle à  $u=0$  avec tangente nulle. Un point d'inflexion à  $u = \gamma - 1 - \sqrt{\gamma - 1}$ , mode à  $u = \gamma - 1$ , second point d'inflexion à  $u = \gamma - 1 + \sqrt{\gamma - 1}$ . Courbe en cloche asymétrique.

Nous montrons en figure 10 les différentes formes des représentations graphiques de la fonction de répartition, et de la fonction de densité d'après les valeurs de  $u$ . Nous donnons également les différentes formes de représentation de  $\log(1 - F)$  en fonction de  $u$ , c'est-à-dire les représentations de  $(1 - F)$  en coordonnées semi-logarithmiques.

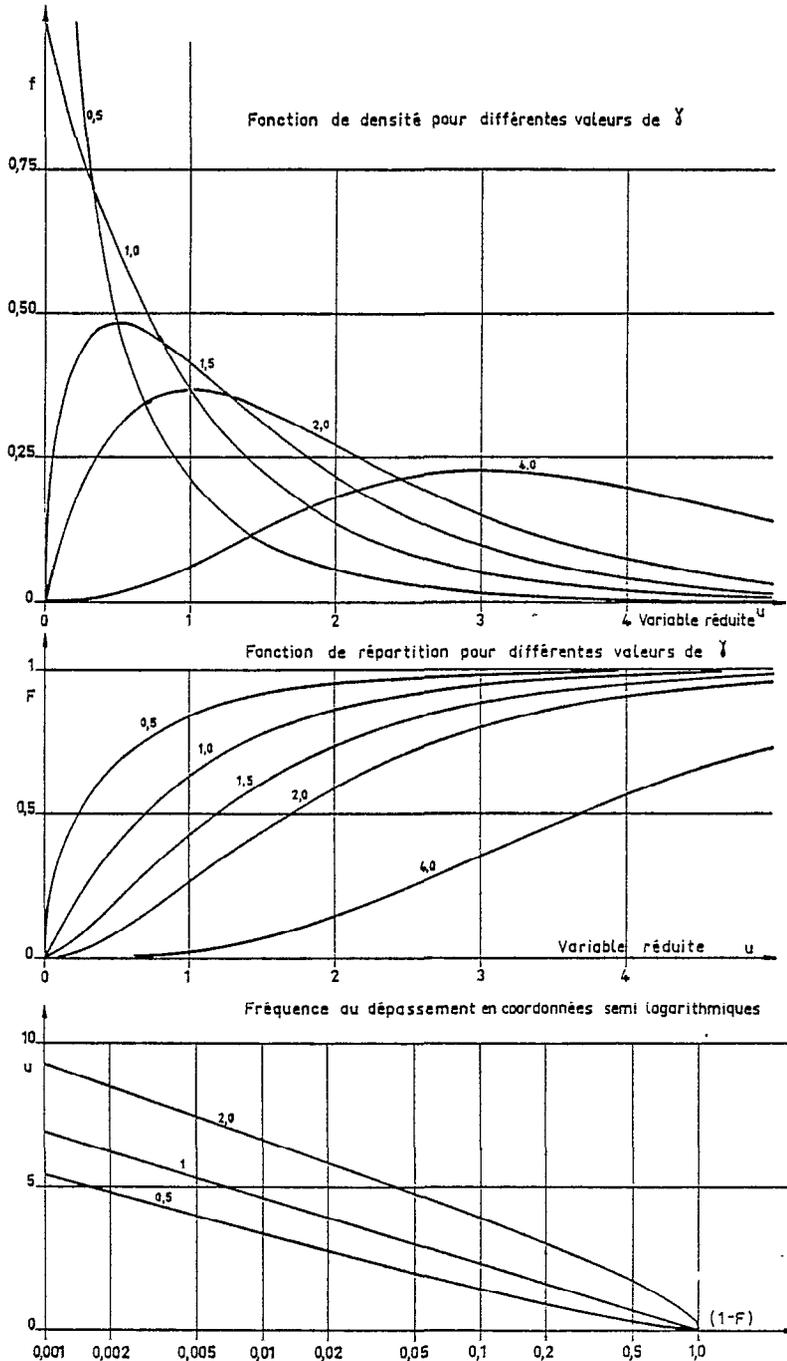


FIG. 10.

7.1.3. — Moments, cumulants, valeurs centrales d'une distribution gamma incomplète parfaitement connue.

Le moment non centré d'ordre  $r$  de la variable réduite est :

$$m_r = \frac{\Gamma(\gamma + r)}{\Gamma(\gamma)} = \gamma(\gamma + 1) \dots (\gamma + r - 1)$$

dont on peut déduire le moment non centré d'ordre  $r$  de la variable non réduite en faisant la transformation  $u = (x - x_0)/s$  (cf paragraphe 1.3.2.). Si le paramètre de position est nul :

$$m_{s,r} = s^r m_r.$$

La variance du moment d'ordre  $r$  de la variable réduite est, pour un échantillon de taille  $n$  tiré d'une distribution gamma incomplète parfaitement connue :

$$\text{var } m_r = \frac{1}{n} (m_{2r} - m_r^2).$$

Le cumulants d'ordre  $r$  de la variable réduite est :

$$k_r = \gamma(r - 1)!$$

et les cumulants de la variable non réduite s'en déduisent par la relation ( $r > 1$ ) :

$$K_{s, x_0, r} = s^r K_r.$$

La moyenne est :  $x_0 + \gamma s$  ;

La variance est :  $K_2 = \gamma s^2$  ;

Le cumulants d'ordre 3 est :  $K_3 = 2 \gamma s^3$  ;

Le cumulants d'ordre 4 est :  $K_4 = 6 \gamma s^4$ .

Le coefficient d'asymétrie est :  $\gamma_1 = 2/\sqrt{\gamma}$  ; l'asymétrie est positive et diminue lorsque  $\gamma$  augmente.

Le coefficient d'aplatissement est :  $\gamma_2 = 6/\gamma$ . Il est toujours positif, et la distribution est d'autant moins aplatie par rapport à la normale que  $\gamma$  est plus petit. Lorsque  $\gamma$  croît indéfiniment, la distribution gamma incomplète tend vers une distribution normale, les deux coefficients  $\gamma_1$  et  $\gamma_2$  tendant vers zéro. (Figure 11.)

Le mode est  $x_0 + (\gamma - 1)s$  ; il n'existe que pour  $\gamma > 1$  ; la médiane est égale à :

$$x_0 + \left(\gamma - \frac{1}{3}\right)s + s \left[ \frac{0,019753}{\gamma} + \frac{0,00721144}{\gamma^2} + \frac{0,00038554}{\gamma^3} + \dots \right].$$

La moyenne harmonique de la variable réduite est :  $(\gamma - 1)$  qui n'a de sens que pour  $\gamma > 1$ .

L'inverse de la moyenne harmonique est :  $1/(\gamma - 1)$  dont la variance est  $\frac{1}{(\gamma - 1)^2(\gamma - 2)}$  (pour  $\gamma > 2$ ).

Si le paramètre de position est nul, la moyenne harmonique dans la loi non réduite est :  $s(\gamma - 1)$ .

La moyenne logarithmique de la variable réduite est :  $\psi(\gamma)$ , dérivée logarithmique de la fonction  $\Gamma(\gamma)$  par rapport à  $\gamma$ .

$\psi(\gamma) = \frac{1}{\Gamma(\gamma)} \int_0^\infty \text{L} u u^{\gamma-1} e^{-u} du$ . La variance de la moyenne logarithmique est  $\psi'(\gamma)$  dérivée de  $\psi(\gamma)$  par rapport à  $\gamma$ .

La moyenne géométrique est égale à  $e^{\psi(\gamma)}$  dans la loi réduite, et si le paramètre de position est nul, elle est égale à  $s e^{\psi(\gamma)}$  dans la loi non réduite.

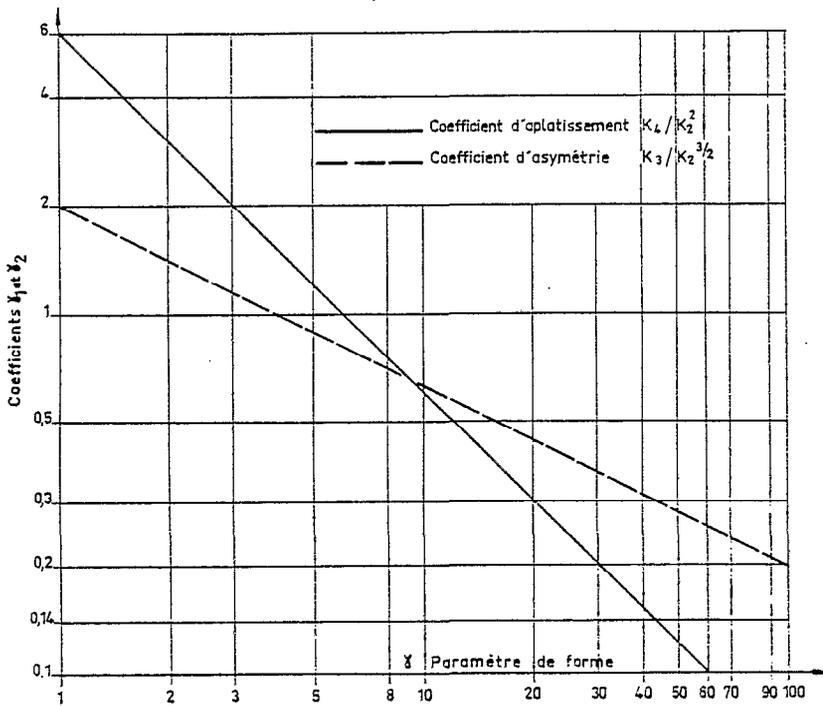


FIG. 11.

La figure 12 montre les positions relatives suivant les valeurs de  $\gamma$  et pour des variables réduites : du mode, de la médiane, et des moyennes arithmétique, géométrique et harmonique.

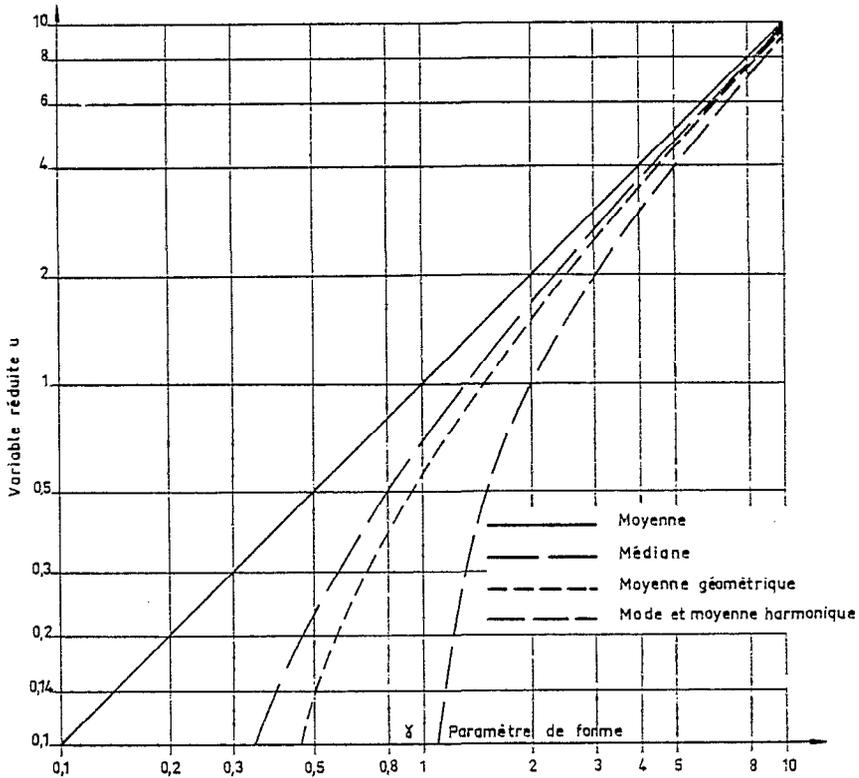


FIG. 12.

Les erreurs-types des valeurs centrales, pour un échantillon de taille  $n$  tiré d'une distribution dont les paramètres  $\gamma$ ,  $s$  et  $x_0$  sont connus par avance, sont les suivantes :

$$\begin{aligned} \text{Erreur type de la moyenne :} & \quad s \sqrt{\frac{\gamma}{n}} \\ \text{Erreur type de la variance :} & \quad s^2 \sqrt{\frac{2 \gamma (\gamma + 3)}{n}}, \\ \text{Erreur type de la médiane :} & \quad 1,2533 s \sqrt{\frac{\gamma}{n}}, \\ \text{Erreur type de l'écart-type :} & \quad s \sqrt{\frac{\gamma + 3}{2 \cdot n}}. \end{aligned}$$

## 7.2. — Estimation des paramètres par les moments (échantillon de taille $n$ ).

En utilisant les sommes des trois premières puissances des valeurs observées ( $x_1 \dots x_i \dots x_n$ ) et en posant :

$$S_1 = \frac{1}{n} \sum x_i, \quad S_2 = \frac{1}{n} \sum x_i^2, \quad S_3 = \frac{1}{n} \sum x_i^3.$$

nous avons les équations :

$$\gamma s - x_0 = S_1,$$

$$\gamma s^2 = \frac{n}{n-1} (S_2 - S_1^2),$$

$$2 \gamma s^3 = \frac{n^2}{(n-1)(n-2)} (S_3 - 3 S_2 S_1 + 2 S_1^3).$$

d'où :

$$s = \frac{n}{2(n-2)} \frac{(S_3 - 3 S_2 S_1 + 2 S_1^3)}{S_2 - S_1^2},$$

$$\gamma = \frac{4(n-2)^2}{n(n-1)} \frac{(S_2 - S_1^2)_3}{(S_3 - 3 S_2 S_1 + 2 S_1^3)^2},$$

$$x_0 = \frac{n-2}{n-1} \frac{(S_2 - S_1^2)^2}{S_3 - 3 S_2 S_1 + 2 S_1^3} - S_1.$$

7.2.1. — Si  $x_0$  a été choisi *a priori*, il est plus simple de passer par les rapports des moments non centrés :

$$R_1 = \frac{\sum (x_i - x_0)}{n} = \gamma s,$$

$$R_2 = \frac{\sum (x_i - x_0)^2}{\sum (x_i - x_0)} = \gamma s + s,$$

$$R_3 = \frac{\sum (x_i - x_0)^3}{\sum (x_i - x_0)^2} = \gamma s + 2 s.$$

Pour un échantillon de taille infinie la quantité :  $R_1 + R_3 - 2 R_2$  est nulle. Appelons résidu la quantité :  $R = R_1 + R_3 - 2 R_2$  qui n'est pas forcément nulle pour un échantillon de taille finie, mais dont la valeur moyenne est nulle pour un grand nombre d'échantillons de même taille

tirés de la même population mère. En supposant linéaire la corrélation qui existe entre le résidu  $R$  et le rapport  $R_2$ , nous pouvons améliorer la valeur brute de  $R_2$  (le coefficient de corrélation de  $R$  avec  $R_1$  est nul).

$$R_{2a} = R_2 - R \frac{2}{3 + 7/(\gamma + 1)}, \text{ la première valeur de } \gamma \text{ ayant été calculée par :}$$

$$\gamma = R_1/(R_2 - R_1).$$

Les valeurs finales des paramètres seront :

$$s = R_{2a} - R_1,$$

$$\gamma = R_1/s.$$

### 7.3. — Détermination des paramètres par le maximum de vraisemblance (échantillon de taille $n$ ).

Pour une valeur  $x_i$  de la variable, la fonction de densité s'écrit :

$$\frac{1}{\Gamma(\gamma)} s^{-\gamma} (x_i - x_0)^{\gamma-1} e^{-(x_i - x_0)/s}.$$

le logarithme naturel du maximum de vraisemblance s'écrit :

$$-n \text{ Log } (\Gamma(\gamma)) - n \gamma \text{ Log } s + (\gamma - 1) \Sigma \text{ Log } (x_i - x_0) - \frac{1}{s} \Sigma (x_i - x_0).$$

En dérivant par rapport à  $x_0$ ,  $s$  et  $\gamma$  et en annulant les dérivées on obtient :

$$(\text{dérivée par rapport à } x_0) \quad n/s - (\gamma - 1) \Sigma \frac{1}{x_i - x_0} = 0,$$

$$\frac{1}{s} \Sigma (x_i - x_0) - n \gamma = 0,$$

$$\Sigma \text{ Log } (x_i - x_0) - n \text{ Log } s - n \psi(\gamma) = 0.$$

La première équation n'est pas utilisable pour  $\gamma \leq 1$  : car la moyenne harmonique n'a pas de sens pour ces valeurs de  $\gamma$ .

En posant :  $\bar{x} = \frac{1}{n} \Sigma x_i$ ,

$$S_h = \frac{1}{n} \Sigma \frac{1}{x - x_0}, \quad S_1 = \frac{1}{n} \Sigma \text{ Log } (x_i - x_0), \quad S_2 = \frac{1}{n} \Sigma x_i^2,$$

et en supposant d'abord  $\gamma > 1$ , il vient :

$$s = (\bar{x} - x_0)/\gamma, \quad \text{donnant } s \text{ en fonction de } x_0 \text{ et de } \gamma,$$

$$\frac{\gamma}{\gamma - 1} = S_h (\bar{x} - x_0) = z,$$

$$\text{Log } \gamma - \psi(\gamma) = \text{Log } (\bar{x} - x_0) - S_1 = y.$$

Nous avons un système de deux équations implicites et transcendantes pour déterminer  $x_0$  et  $\gamma$ . La méthode qui semble la meilleure à utiliser est de choisir une valeur de  $x_0$ , calculer  $z$  et  $y$  d'où des valeurs pour  $\gamma$  :

$$\begin{aligned}\gamma_z &= z/(z-1), \\ \gamma_y &= 0,5000876/y + 0,1648852 - 0,0544274 y.\end{aligned}$$

cette dernière formule représentant une solution approchée (avec une erreur relative maximale sur  $\gamma_y$  de 0,01 %) de l'équation :

$$\text{Log } \gamma - \psi(\gamma) = y \quad \text{pour } \gamma \geq 1$$

(qui pourrait être résolue en  $\gamma$  par une des méthodes d'inversion de fonctions).

On compare  $\gamma_z$  et  $\gamma_y$  pour savoir dans quel sens faire varier  $x_0$  : si  $(\gamma_z - \gamma_y)$  est positif, la valeur choisie pour  $x_0$  est trop petite, si  $(\gamma_z - \gamma_y)$  est négatif, la valeur choisie pour  $x_0$  est trop grande.

D'où le choix d'une nouvelle valeur de  $x_0$  et un nouveau calcul de  $\gamma_z$  et de  $\gamma_y$ . L'on procède par approximations successives jusqu'à trouver :  $\gamma_z/\gamma_y = 1$  à 0,01 % près.

En supposant  $\gamma \leq 1$  nous suivons une marche analogue à celle décrite ci-dessus, avec :

$$\gamma_z = \frac{n-1}{n} (\bar{x} - x_0)^2 / (S_2 - \bar{x}^2),$$

en remplaçant l'équation contenant la moyenne harmonique par l'équation calculant le moment centré de second ordre, et :

$$\gamma_y = (8,898919/y + 9,059950 + 0,9775373 y)/(17,79728 + 11,968477 y + y^2),$$

formule qui représente une solution approchée de l'équation :

$$\text{Log } \gamma - \psi(\gamma) = y \quad \text{pour } \gamma \leq 1.$$

La comparaison de  $\gamma_z$  et de  $\gamma_y$  donne une règle inverse de celle trouvée ci-dessus : si  $(\gamma_z - \gamma_y)$  est positif, la valeur choisie pour  $x_0$  est trop grande...

#### 7.4. — Estimation des paramètres : loi gamma incomplète tronquée avec troncature. Estimation par les moments.

Nous nous plaçons dans le cas de lois en « J » ( $\gamma \leq 1$ ) ou dans le cas de lois dont la fonction de densité admet une tangente infinie à l'origine ( $1 < \gamma < 2$ ), ces lois étant tronquées (cf. paragraphes 1.4. et 2.5.3.), c'est-à-dire que la fréquence des observations inférieures ou égales à la borne inférieure  $x_0$  n'est pas nulle. De plus, nous nous plaçons dans le cas d'une troncature, c'est-à-dire que le nombre des observations qui auraient dû être faites dans la tranche  $x_0, x_h$  ( $x_h$  seuil de troncature, supérieur à la borne  $x_0$ ) est inconnu, ainsi que les valeurs individuelles de ces observations.

Le seuil de troncature peut être déterminé par la sensibilité de l'appareil de mesure, ou ses conditions d'emploi. Prenons par exemple un échantillon de hauteurs pluviométriques journalières : on peut choisir le seuil de troncature à 0,1 mm puisque les quantités inférieures à cette hauteur ne sont pas mesurées ; mais il vaut mieux choisir un seuil plus élevé (5 ou 10 mm) car l'évaporation dans le pluviomètre n'est pas négligeable pendant les journées à faibles hauteurs pluviométriques qui seront connues par défaut quels que soient la conscience et le soin de l'observateur.

Nous pouvons admettre que la borne inférieure  $x_0$  est connue par avance avec une approximation suffisante. En posant :

$$\begin{aligned} S_1 &= \Sigma (x_i - x_0), & S_2 &= \Sigma (x_i - x_0)^2, & S_3 &= \Sigma (x_i - x_0)^3, \\ & & S_4 &= \Sigma (x_i - x_0)^4, \end{aligned}$$

sommes qui sont convenablement connues car les observations non faites ( $x_i$  peu différent de  $x_0$ ) et donc non utilisées interviendraient pour peu de chose dans ces totaux.

Nous écrirons les équations :

$$\begin{aligned} R_2 &= S_2/S_1 = \gamma s + s, \\ R_3 &= S_3/S_2 = \gamma s + 2s, \\ R_4 &= S_4/S_3 = \gamma s + 3s. \end{aligned}$$

Il suffit en principe des deux premières pour calculer les paramètres  $\gamma$  et  $s$  :

$$\begin{aligned} s &= R_3 - R_2, \\ \gamma &= \frac{1}{s} (2R_2 - R_3). \end{aligned}$$

Le nombre « théorique » d'observations est  $M$  donné par :

$$M = \frac{1}{s \gamma} S_1,$$

et si l'on appelle  $A$  le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, le paramètre  $F_0$  est calculé par :

$$F_0 = \frac{A - M}{A}.$$

On peut définir comme plus haut (paragraphe 7.2.1.) un résidu  $R = R_2 + R_4 - 2 R_3$ , de valeur moyenne nulle. En supposant linéaire la corrélation qui existe entre le résidu  $R$  et les rapports  $R_2$  et  $R_3$ , nous pouvons améliorer les valeurs brutes de ces rapports :

$$\begin{aligned} R_{2a} &= R_2 - R \times 2 / [3 + 37/(\gamma + 1) + 39/(\gamma + 1)(\gamma + 2)], \\ R_{3a} &= R_3 - R \times [7 + 16/(\gamma + 1)] / [3 + 37/(\gamma + 1) + 39/(\gamma + 1)(\gamma + 2)], \end{aligned}$$

la première valeur de  $\gamma$  ayant été calculée par :  $\gamma = \frac{2 R_2 - R_3}{R_3 - R_2}$ ,

les valeurs finales des paramètres seront :

$$\begin{aligned} s &= R_{3a} - R_{2a}, \\ \gamma &= \frac{2 R_{2a} - R_{3a}}{R_{3a} - R_{2a}}, \\ F_0 &= 1 - \frac{S_1}{A (2 R_{2a} - R_{3a})}. \end{aligned}$$

7.5. — Détermination des paramètres par le maximum de vraisemblance : loi gamma tronquée avec troncature : la borne inférieure  $x_0$  étant connue par avance, et le seuil de troncature  $x_h$  étant choisi (cf. paragraphe précédent).

On applique la méthode aux  $n$  observations conservées ( $x \geq x_h$ ) dont la probabilité totale est :

$$\frac{1 - F_0}{\Gamma(\gamma)} \int_{x_h}^{\infty} u^{\gamma-1} e^{-u} du \quad \text{avec } u = (x - x_0)/s.$$

Pour une valeur  $x_i$  de la variable la fonction de densité s'écrit :

$$\frac{(1 - F_0)}{\Gamma(\gamma)} s^{-\gamma} (x_i - x_0)^{\gamma-1} e^{-(x_i - x_0)/s} \left/ \frac{(1 - F_0)}{\Gamma(\gamma)} \int_{x_h}^{\infty} u^{\gamma-1} e^{-u} du \right.$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$(\gamma - 1) \Sigma \text{Log}(x_i - x_0) - n \gamma \text{Log } s - \frac{1}{s} \Sigma (x_i - x_0) - n \text{Log} \left[ \Gamma(\gamma) - \int_{x_0}^{x_h} u^{\gamma-1} e^{-u} du. \right]$$

En dérivant par rapport à  $s$  et  $\gamma$  et en annulant les dérivées on obtient :

$$-\gamma + \frac{1}{s} \frac{\Sigma (x_i - x_0)}{n} - \left( \frac{x_h - x_0}{s} \right)^{\gamma} e^{-\frac{x_h - x_0}{s}} \left/ \left[ \Gamma(\gamma) - \int_{x_0}^{x_h} u^{\gamma-1} e^{-u} du \right] \right. = 0,$$

$$\text{Log } s - \frac{\Sigma \text{Log}(x_i - x_0)}{n} + \left[ \psi(\gamma) \Gamma(\gamma) - \int_{x_0}^{x_h} Lu u^{\gamma-1} e^{-u} du \right] \left/ \left[ \Gamma(\gamma) - \int_{x_0}^{x_h} u^{\gamma-1} e^{-u} du \right] \right. = 0.$$

en posant :  $u_h = \frac{x_h - x_0}{s}$  et  $P_h = \frac{1}{\Gamma(\gamma)} \int_0^{u_h} u^{\gamma-1} e^{-u} du,$

$$\gamma = \frac{\Sigma (x_i - x_0)}{n s} - u_h^{\gamma} e^{-u_h} / [\Gamma(\gamma) (1 - P_h)],$$

$$\text{Log } s = \frac{\Sigma \text{Log}(x_i - x_0)}{n} + \left[ -\psi(\gamma) \Gamma(\gamma) + \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \frac{u_h^{\gamma+r}}{\gamma+r} \left( \text{Log } u_h - \frac{1}{\gamma+r} \right) \right] \left/ \left[ \Gamma(\gamma) (1 - P_h) \right] \right.$$

La série  $\sum_{r=0}^{\infty}$  est alternée, absolument convergente, et très rapidement convergente si  $u_h < 1$ .

Nous nous trouvons en présence d'un système de deux équations transcendantes et implicites en  $s$  et  $\gamma$ . Il est impossible de choisir des valeurs approchées de  $\gamma$  et de  $s$  à porter dans la pre-

mière (ou la deuxième) de ces équations pour calculer une valeur de  $\gamma$  (ou de  $\text{Log } s$ ) : le système est toujours divergent.

La méthode que nous avons mise au point pour le calcul par ordinateur consiste à choisir au départ un  $\gamma$  calculé par la méthode des moments, puis à déterminer  $s$  par approximations successives pour que :

$$\frac{\sum (x_i - x_0)}{n} \neq \gamma s + s u_h^\gamma e^{-u_h} / [\Gamma(\gamma) (1 - P_h)],$$

on détermine alors  $\gamma$  par approximations successives pour que  $\frac{\sum \text{Log } (x_i - x_0)}{n} \neq \text{Log } s$

$$+ \left[ \psi(\gamma) \Gamma(\gamma) - \sum_{r=0}^{\infty} \frac{(-1)^r u_h^{\gamma+r}}{r! \gamma+r} \left( \text{Log } u_h - \frac{1}{\gamma+r} \right) \right] / \Gamma(\gamma) (1 - P_h),$$

on retourne à la détermination de  $s$ , puis à celle de  $\gamma$  et ainsi de suite. La convergence des déterminations successives de  $s$  et de  $\gamma$  est assez rapide.

Le paramètre de tronquage  $F_0$  est calculé par :

$$F_0 = 1 - \frac{n/A}{1 - P_h}.$$

$A$  étant le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, et  $P_h$  étant la dernière valeur calculée plus haut de  $\frac{1}{\Gamma(\gamma)} \int_0^{u_h} u^{\gamma-1} e^{-u} du$ ,  $n$  étant le nombre d'observations conservées, de valeurs égales ou supérieures au seuil de tronquature.

## 7.6. — Comportement asymptotique de la distribution gamma incomplète.

Si la fréquence  $F$  tend vers 1, et en posant  $T = \frac{1}{1-F}$ ,  $T$  étant le temps de récurrence ou période de retour, la variable réduite  $u$  tend vers  $\text{Log } T - \text{Log } (\Gamma(\gamma)) + (\gamma - 1) \text{Log } u$ .

Il n'y a pas d'asymptote à proprement parler, mais on peut dire que  $u$  tend vers  $\text{Log } T$  d'une façon parabolique.

## 7.7. — Calculs à l'ordinateur.

7.7.1. — Calcul de la probabilité correspondant à une valeur donnée de la variable.

Il s'agit d'utiliser un développement limité permettant de calculer :

$$F(x) = \frac{1}{\Gamma(\gamma)} \int_0^u u^{\gamma-1} e^{-u} du \quad \text{avec} \quad u = \frac{(x - x_0)}{s}.$$

Cette expression peut s'intégrer par partie de deux façons différentes :

soit, en posant  $\gamma = \delta + r$  où  $r$  est entier positif et  $\delta > 0$  :

$$F(x) = -e^{-u} \sum_{i=1}^r \frac{u^{\delta+i-1}}{\Gamma(\delta+i)} + \frac{1}{\Gamma(\delta)} \int_0^u u^{\delta-1} e^{-u} du.$$

les  $r$  termes du développement limité se déduisant l'un de l'autre par :  $U_{i+1} = \frac{u}{\delta+i} U_i$ ,

soit, en posant  $\delta = \gamma + r$  où  $r$  est entier positif :

$$F(x) = e^{-u} \sum_{i=1}^r \frac{u^{\gamma+i-1}}{\Gamma(\gamma+i)} + \frac{1}{\Gamma(\delta)} \int_0^u u^{\delta-1} e^{-u} du,$$

les  $r$  termes du développement limité se déduisant l'un de l'autre par  $U_{i+1} = \frac{u}{\gamma+i} U_i$

Lorsque  $u$  est supérieur ou égal à 17,0, nous utilisons le premier développement avec  $r$  termes, ce nombre étant calculé pour que :  $\frac{1}{\Gamma(\delta)} \int_0^u u^{\delta-1} e^{-u} du$  soit égal à 1 avec une erreur d'au plus 0,00000005, ce qui est atteint pour :

$$\delta \leq (\sqrt{u} - 2,81)^2 - 0,7,$$

$r$  est alors la partie entière de l'expression  $\gamma - \delta + 1$ .

Lorsque  $u$  est inférieur à 17,0 nous utilisons le second développement avec  $r$  termes, ce nombre étant calculé pour que :  $\frac{1}{\Gamma(\delta)} \int_0^u u^{\delta-1} e^{-u} du$ , soit égal à zéro avec une erreur d'au plus 0,00000005, ce qui est atteint pour :

$$\delta \geq (\sqrt{u} + 2,87)^2 - 4,5,$$

$r$  est alors la partie entière de l'expression  $\delta - \gamma$ .

Le programme que nous avons écrit fonctionne avec des valeurs de  $\gamma$  inférieures à 113,0.

7.7.2. — Calcul de la valeur de la variable correspondant à une probabilité donnée.

L'inversion de la fonction gamma incomplète se fait facilement par la méthode générale « du pas » en partant de la valeur moyenne de la variable réduite  $u = \gamma$ .

### 7.8. — Approximations intéressantes.

On se ramène à l'utilisation d'une table de la loi normale en considérant que :  $\sqrt[3]{\frac{x_i - x_0}{x - x_0}}$  est distribué normalement avec pour moyenne  $1 - \frac{1}{9\gamma}$  et pour variance  $\frac{1}{9\gamma}$ . L'approximation est d'autant meilleure que  $\gamma$  est grand. Dès  $\gamma = 8$  l'erreur est inférieure à 1 % sur les probabilités au non dépassement allant de 0,01 à 0,999 et même inférieure à 1 ‰ sur des probabilités au non dépassement allant de 0,05 à 0,99.

Lorsque  $\gamma$  est supérieur à 50 on peut considérer que la quantité :  $\sqrt{4(x_i - x_0)/s} - \sqrt{4\gamma - 1}$ , est distribuée normalement avec, comme moyenne zéro et un comme variance.

Si  $2\gamma$  est un entier, on peut utiliser des tables de  $\chi^2$  en prenant  $2\gamma$  comme nombre de degrés de liberté et  $2 \frac{x_i - x_0}{s}$  comme argument.

La somme des  $\gamma$  premiers termes d'une distribution de Poisson :

$$\sum_{i=0}^{\gamma-1} e^{-m} \frac{m^i}{i!} \quad (\gamma \text{ entier}) \text{ est égale à } 1 - \int_0^m u^{\gamma-1} e^{-u} du.$$

## VIII. — DISTRIBUTION EXPONENTIELLE GÉNÉRALISÉE

### 8.1. — Expressions mathématiques.

#### 8.1.1. — Fonction de répartition.

Nous l'écrivons :  $F(x) = \frac{\sigma + 1}{2} - \sigma e^{-u^{1/\delta}}$ , avec :  $u = \frac{x - x_0}{s}$ ,  $\sigma$  étant égal à 1, avec le signe de  $s \delta$ .

$F(x)$  est la fréquence au non dépassement : lorsque,  $s$  étant positif  $x$  varie de  $x_0$  à  $+\infty$ , ou que,  $s$  étant négatif  $x$  varie de  $-\infty$  à  $x_0$ ,  $F(x)$  croît de 0 à 1 ;

$x_0$  est le paramètre de position, borne inférieure de l'intervalle de définition de la variée si  $s$  est positif, ou borne supérieure si  $s$  est négatif ;

$s$  est le paramètre d'échelle, différent de zéro, ayant les mêmes dimensions que  $x_0$  et  $x$  ;

$\delta$  est le paramètre de forme, différent de zéro.

La distribution exponentielle généralisée correspond :

Si  $s$  et  $\delta$  sont positifs : à la loi de Goodrich, habituellement écrite :

$$F(x) = 1 - e^{-a(x-x_0)^{1/n}} \quad [n = \delta, \quad a = 1/s],$$

si  $s$  est positif et  $\delta$  négatif : à la loi de Fréchet, habituellement écrite :

$$F(x) = e^{-a(x-x_0)^{-k}} \quad [k = -1/\delta, \quad a = 1/s],$$

si le produit  $s \delta$  est négatif : aux lois de Jenkinson, habituellement écrites :

$$F(x) = e^{-\left(1 - \frac{x-x_0}{a}\right)^{1/k}} \quad [k = \delta, \quad a = -s, \quad x' + a = x_0].$$

#### 8.1.2. — La fonction de densité est :

$$f(u) = \frac{1}{|\delta|} u^{\frac{1-\delta}{\delta}} e^{-u^{1/\delta}},$$

dont la dérivée première s'écrit :

$$\frac{1}{\delta^2} u^{\frac{1}{\delta}-2} e^{-u^{1/\delta}} [(1-\delta) - u^{1/\delta}] \quad \text{de racine } u = (1-\delta)^\delta.$$

et la dérivée seconde :

$$\frac{1}{|\delta|^3} u^{\frac{1}{\delta}-3} e^{-u^{1/\delta}} [(1-\delta)(1-2\delta) - 3(1-\delta)u^{1/\delta} + u^{2/\delta}]$$

$$\text{de racines } u = \left[ \frac{3(1-\delta) \pm \sqrt{(1-\delta)(5-\delta)}}{2} \right]^\delta$$

d'où les différentes formes possibles de la représentation graphique de la fonction de densité.

$\delta < 1/2$  (y compris  $\delta < 0$ ), densité de probabilité nulle à  $u = 0$ , deux points d'inflexion encadrant le mode :  $u = (1-\delta)^\delta$ , la densité de probabilité tend vers zéro si  $u$  tend vers  $+\infty$  (Courbe en « cloche ») ;

- $\delta = 1/2$  densité de probabilité nulle à  $u = 0$ , avec tangente  $+ 2$ , mode à  $u = 1/\sqrt{2}$  et un seul point d'inflexion à  $u = 3/2$ .
- $\frac{1}{2} < \delta < 1$  densité de probabilité nulle à  $u = 0$ , avec tangente  $+ \infty$ , mode à  $u = (1 - \delta)^\delta$  et un point d'inflexion à

$$u = \left[ \frac{3(1 - \delta) + \sqrt{(1 - \delta)(5 - \delta)}}{2} \right]^\delta,$$

- $\delta = 1$  densité de probabilité 1 à  $u = 0$ , pas de mode réel ni de points d'inflexion (mode observé à  $u = 0$ ). Loi exponentielle simple ;
- $1 < \delta$  densité de probabilité infinie à  $u = 0$ , pas de mode réel ni de points d'inflexion (mode observé à  $u = 0$ ). Loi dite en « J ».

Si  $\delta$  tend vers zéro, par valeurs positives ou négatives, la loi exponentielle généralisée tend vers une distribution de Gumbel dont le paramètre d'échelle serait de signe contraire à celui de  $\delta$ .

Nous montrons en figure 13 les différentes formes des représentations graphiques de la fonction de répartition et de la fonction de densité d'après les valeurs de  $u$ . Nous donnons également les différentes formes des représentations de  $\log(1 - F)$  en fonction de  $u$ , c'est-à-dire les représentations de  $(1 - F)$  en coordonnées semi-logarithmiques.

8.1.3. — Moments, cumulants, valeurs centrales d'une distribution exponentielle généralisée parfaitement connue.

Le moment non centré d'ordre  $r$  de la variable réduite est :

$$m_r = \Gamma(\delta r + 1)$$

qui n'existe que pour  $\delta r + 1 > 0$ . Si  $\delta$  est positif, tous les moments existent, mais si  $\delta$  est négatif ils n'existent pas tous. La moyenne elle-même n'a plus de sens pour  $\delta \leq -1$ .

De la formule précédente, on peut déduire le moment non centré d'ordre  $r \geq 1$  de la variable non réduite en faisant la transformation  $u = \frac{x - x_0}{s}$  (cf. paragraphe 1.3.2.). Si le paramètre de position est nul :

$$m_{sr} = s^r m_r.$$

La variance du moment d'ordre  $r$  de la variable réduite est, pour un échantillon de taille  $n$  tiré d'une distribution exponentielle généralisée parfaitement connue :

$$\text{var } m_r = \frac{1}{n} (m_{2r} - m_r^2).$$

La moyenne est :

$$x_0 + s \Gamma(\delta + 1) \text{ pour } \delta > -1;$$

$$K_2 = s^2 [\Gamma(2\delta + 1) - (\Gamma(\delta + 1))^2] \text{ pour } \delta > -1/2;$$

$$K_3 = s^3 [\Gamma(3\delta + 1) - 3 \Gamma(2\delta + 1) \Gamma(\delta + 1) + 2 (\Gamma(\delta + 1))^3] \text{ pour } \delta > -1/3;$$

$$K_4 = s^4 [\Gamma(4\delta + 1) - 4 \Gamma(3\delta + 1) \Gamma(\delta + 1) - 3 (\Gamma(2\delta + 1))^2 + 12 \Gamma(2\delta + 1) (\Gamma(\delta + 1))^2 - 6 (\Gamma(\delta + 1))^4] \text{ pour } \delta > -1/4.$$

Le coefficient d'asymétrie  $\gamma_1$  et le coefficient d'aplatissement  $\gamma_2$  sont d'expressions compliquées : le premier n'a plus de sens pour  $\delta < -1/3$  et le second pour  $\delta < -1/4$ . Le graphique 14 montre les variations de ces coefficients suivant  $\delta$ .

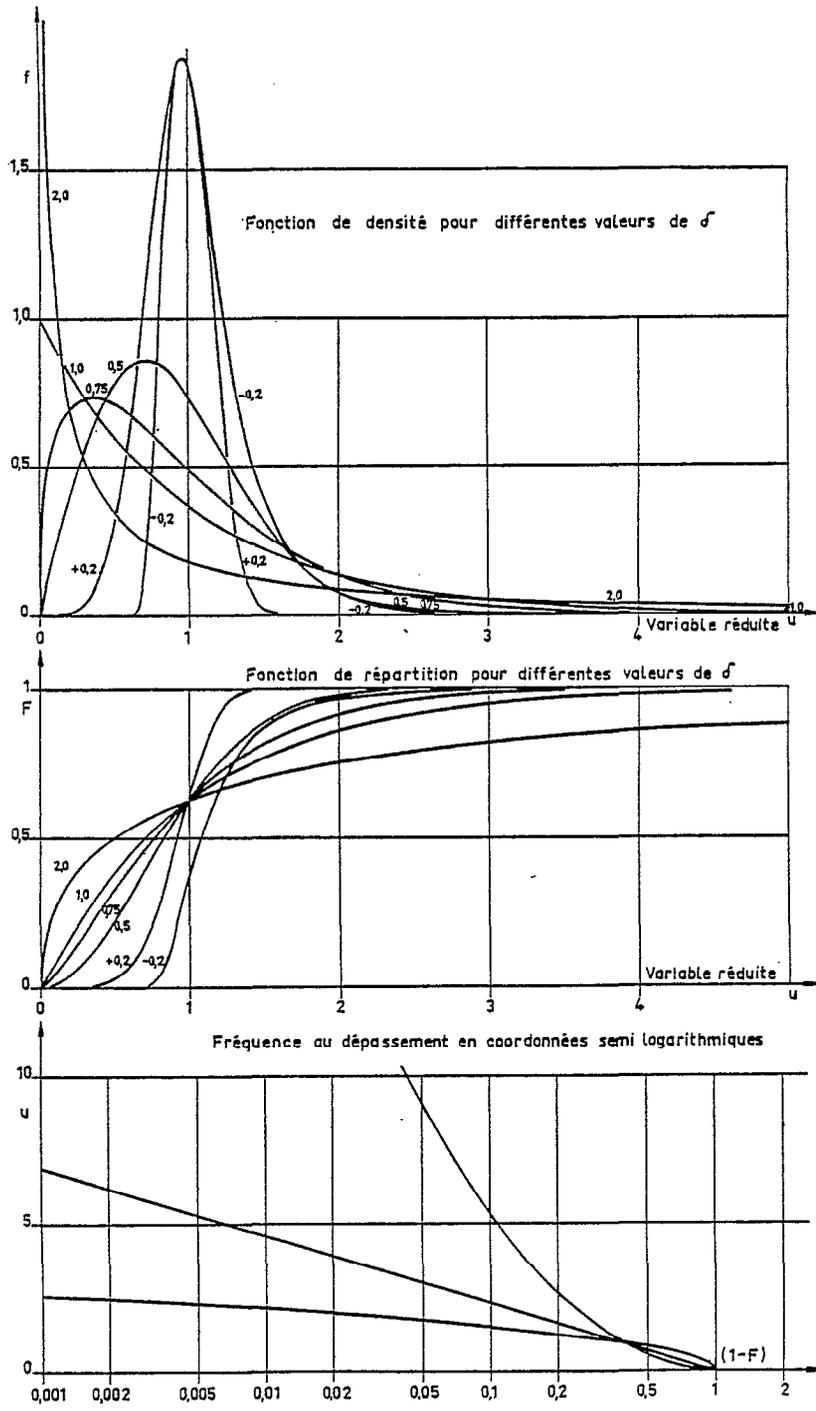


FIG. 13.

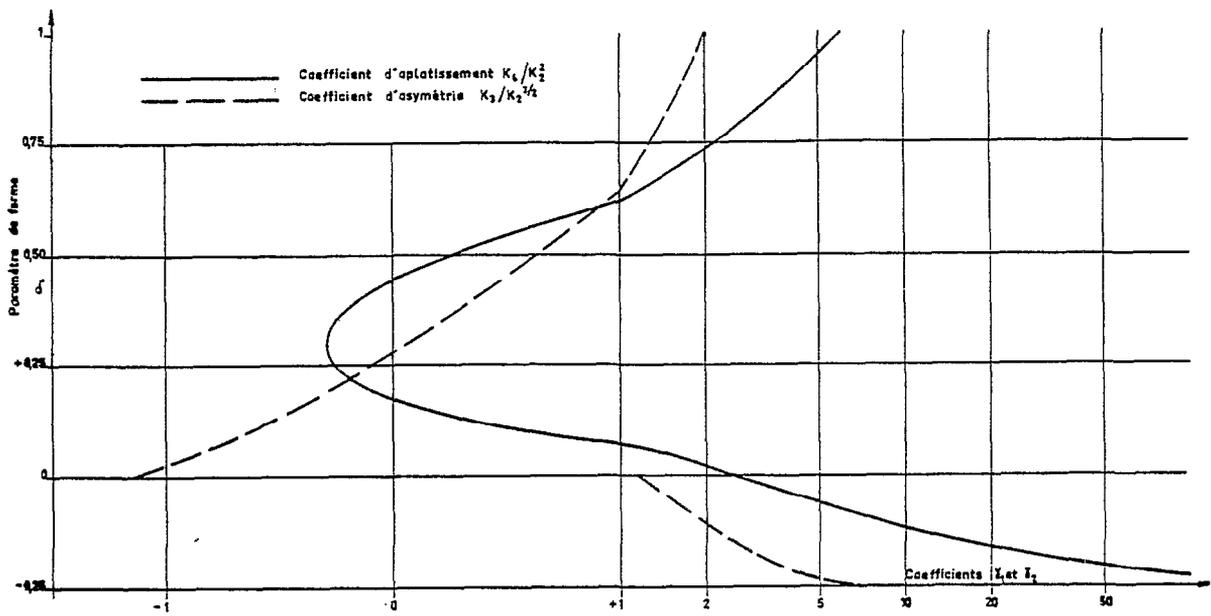


FIG. 14.

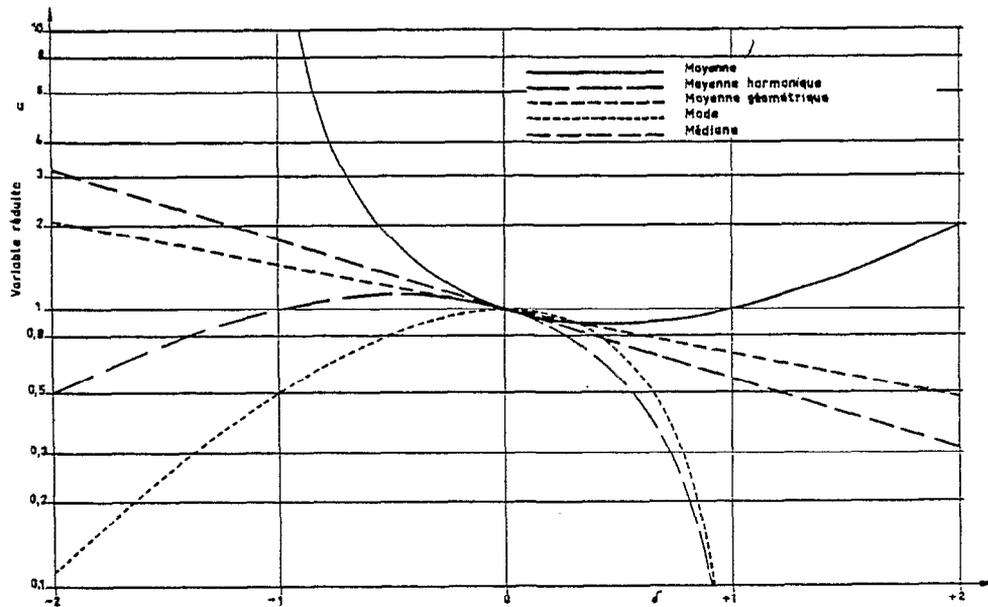


FIG. 15.

$-1/4 < \delta < 0 = \gamma_1$  positif, distribution (de la variable réduite) étalée sur la droite;  
 $\gamma_2$  positif, distribution moins aplatie que la normale;  
 $0 < \delta < 0,173 \quad \gamma_1$  négatif, distribution étalée sur la gauche;  
 $\gamma_2$  positif;  
 $0,173 < \delta < 0,28 \quad \gamma_1$  négatif;  
 $\gamma_2$  négatif, distribution plus aplatie que la normale;  
 $0,28 < \delta < 0,443 \quad \gamma_1$  positif, distribution étalée sur la droite;  
 $\gamma_2$  négatif;  
 $0,443 < \delta \quad \gamma_1$  positif;  
 $\gamma_2$  positif, distribution moins aplatie que la normale.

Le mode,  $x_0 + s(1 - \delta)^\delta$ , n'existe que pour  $\delta < 1$ , la médiane est  $x_0 + s(0,693147)^\delta$ , la moyenne harmonique de la variable réduite est  $1/\Gamma(1 - \delta)$ , qui n'a de sens que pour  $\delta < 1$ . L'inverse de la moyenne harmonique est  $\Gamma(1 - \delta)$  dont la variance  $\Gamma(1 - 2\delta) - (\Gamma(1 - \delta))^2$  n'a de sens que pour  $\delta < 1/2$ . Si le paramètre de position est nul, la moyenne harmonique dans la loi non réduite est  $s/\Gamma(1 - \delta)$ .

La moyenne logarithmique de la variable réduite est de  $-\delta C$  (avec  $C = 0,5772 \dots$ ) et sa variance est de  $\delta^2 \frac{\pi^2}{6}$ . La moyenne géométrique est  $e^{-\delta C} = (0,56142 \dots)^\delta$  pour la variable réduite, et, si le paramètre de position est nul, de  $s(0,56142 \dots)^\delta$  pour la variable non réduite.

La figure 15 montre les positions relatives, suivant les valeurs de  $\delta$  et pour des variables réduites : du mode, de la médiane et des moyennes arithmétique, géométrique et harmonique.

## 8.2. — Estimation des paramètres par les moments (échantillon de taille $n$ ).

Il faut choisir *a priori* le signe  $s_i$  du paramètre d'échelle  $s$  suivant que l'on désire une borne inférieure ( $s > 0$ ,  $s_i = +1$ ) ou supérieure ( $s < 0$ ,  $s_i = -1$ ) de l'intervalle de définition de la variable. Le calcul n'est faisable que pour  $\delta > -1/3$ .

En utilisant les sommes des trois premières puissances des valeurs observées ( $x_1 \dots x_i \dots x_n$ ) et en posant :

$$S_1 = \frac{1}{n} \sum x_i, \quad S_2 = \frac{1}{n} \sum x_i^2, \quad S_3 = \frac{1}{n} \sum x_i^3,$$

nous avons les équations :

$$\begin{aligned}
 s \Gamma(\delta + 1) + x_0 &= S_1, \\
 s^2 \left[ \Gamma(2\delta + 1) - (\Gamma(\delta + 1))^2 \right] &= \frac{n}{n-1} (S_2 - S_1^2), \\
 s^3 \left[ \Gamma(3\delta + 1) - 3 \Gamma(\delta + 1) \cdot \Gamma(2\delta + 1) + 2 (\Gamma(\delta + 1))^3 \right] \\
 &= \frac{n^2}{(n-1)(n-2)} (S_3 - 3 S_2 S_1 + 2 S_1^3).
 \end{aligned}$$

On élimine le paramètre d'échelle en passant par le coefficient d'asymétrie  $K_3/K_2^{3/2}$

$$\begin{aligned}
 \frac{\sqrt{n(n-1)}}{n-2} \frac{S_3 - 3 S_2 S_1 + 2 S_1^3}{(S_2 - S_1^2) \sqrt{(S_2 - S_1^2)}} &= \frac{s_i}{\sqrt{\Gamma(2\delta + 1) - (\Gamma(\delta + 1))^2}} \\
 &\times \left[ \frac{\Gamma(3\delta + 1) - (\Gamma(\delta + 1))^3}{\Gamma(2\delta + 1) - (\Gamma(\delta + 1))^2} - 3 \Gamma(\delta + 1) \right],
 \end{aligned}$$

équation qui permet de déterminer  $\delta$  par approximations successives.

$s$  est donné par :

$$s = s_i \sqrt{\frac{n}{n-1}} \sqrt{\frac{S_2 - S_1^2}{\Gamma(2\delta + 1) - (\Gamma(\delta + 1))^2}},$$

et  $x_0$  par :

$$x_0 = S_1 - s \Gamma(\delta + 1).$$

8.2.1. — Si  $x_0$  a été choisi *a priori* ainsi que le signe  $s_i$  de  $s$  il est plus simple de passer par les rapports des moments non centrés :

$$R_1 = \frac{\Sigma (x_i - x_0)}{n} = s \Gamma(\delta + 1),$$

$$R_2 = \frac{\Sigma (x_i - x_0)^2}{\Sigma (x_i - x_0)} = s \frac{\Gamma(2\delta + 1)}{\Gamma(\delta + 1)},$$

d'où l'équation déterminant  $\delta$  par approximations successives :

$$\frac{\Gamma(2\delta + 1)}{\Gamma(\delta^2 + 1)} = \frac{R_2}{R_1}, \quad \delta \text{ doit être } > -1/2,$$

et

$$s = R_1 / \Gamma(\delta + 1).$$

### 8.3. — Détermination des paramètres par le maximum de vraisemblance (échantillon de taille $n$ ).

Pour une valeur  $x_i$  de la variable, la fonction de densité s'écrit :

$$\frac{1}{|s \delta|} \left( \frac{x_i - x_0}{s} \right)^{\frac{1}{\delta} - 1} e^{-\left( \frac{x_i - x_0}{s} \right)^{1/\delta}},$$

le logarithme naturel du maximum de vraisemblance s'écrit :

$$-n \text{Log} |\delta| - \frac{n}{\delta} \text{Log} |s| + \left( \frac{1}{\delta} - 1 \right) \Sigma \text{Log} |x_i - x_0| - \Sigma |x_i - x_0|^{1/\delta} |s|^{-1/\delta},$$

en dérivant par rapport à  $x_0$ ,  $s$  et  $\delta$  et en annulant les dérivées on obtient :

$$(\text{dérivée par rapport à } x_0) - \left( \frac{1}{\delta} - 1 \right) \Sigma \frac{1}{|x_i - x_0|} + |s|^{-\frac{1}{\delta}} \frac{1}{\delta} \Sigma |x_i - x_0|^{1/\delta - 1} = 0,$$

$$|s|^{-1/\delta} \Sigma |x_i - x_0|^{1/\delta} = n,$$

$$|\delta| = \frac{\Sigma |x_i - x_0|^{1/\delta} \text{Log} |x_i - x_0|}{\Sigma |x_i - x_0|^{1/\delta}} - \frac{1}{n} \Sigma \text{Log} |x_i - x_0|.$$

La première équation n'est pas utilisable pour  $\delta \geq 1$ , car la moyenne harmonique n'a plus de sens pour ces valeurs de  $\delta$ . Nous pouvons remplacer cette équation, soit par une formule basée sur la moyenne arithmétique, soit par une formule basée sur la moyenne logarithmique (dont le coefficient de variation est inférieur à celui de la moyenne arithmétique pour  $\delta > 2$ ), c'est-à-dire :

$$\frac{\Sigma x_i - x_0}{s} = n \Gamma(\delta + 1),$$

ou :

$$\Sigma \text{Log} |x_i - x_0| = n (\text{Log} |s| - \delta \cdot C).$$

En posant :  $C = 0,5772156649$ .

$$SD = \frac{1}{n} \Sigma |x_i - x_0|^{1/\delta}, \quad SL = \frac{1}{n} \Sigma \text{Log} |x_i - x_0|, \quad SM = \frac{1}{n} \Sigma |x_i - x_0|^{1/\delta} \text{Log} |x_i - x_0|,$$

$$SH = \frac{1}{n} \Sigma \left| \frac{1}{x_i - x_0} \right|, \quad SE = \frac{1}{n} \Sigma |x_i - x_0|^{\frac{1}{\delta} - 1},$$

et en supposant d'abord  $\delta < 1$ , il vient :

$$\begin{aligned} |s| &= SD^\delta, \\ \delta &= SM/SD - SL, \\ \delta &= 1 - SE/(SH - SD). \end{aligned}$$

Nous avons un système de deux équations implicites et transcendantes pour déterminer  $x_0$  et  $\delta$ . La méthode qui semble la meilleure est de choisir d'abord le signe de  $s$  (positif si l'on désire une borne  $x_0$  inférieure, négatif si l'on désire une borne supérieure), de choisir une valeur de  $x_0$  et de calculer :

$$\begin{aligned} \delta_y &= 1 - SE/(SH \cdot SD), \\ \delta_e &= SM/SD - SL. \end{aligned}$$

On compare  $\delta_y$  et  $\delta_e$  pour savoir dans quel sens il faut faire varier  $x_0$ .

Si  $(\delta_y - \delta_e)$  est positif,  $x_0$  doit varier en sens inverse du signe de  $s$ , si  $(\delta_y - \delta_e)$  est négatif,  $x_0$  doit varier dans le sens du signe de  $s$ . D'où le choix d'une nouvelle valeur de  $x_0$  et un nouveau calcul de  $\delta_y$  et  $\delta_e$ . L'on procède par approximations successives jusqu'à trouver  $\delta_y/\delta_e = 1$ , avec l'erreur relative acceptée.

Comme il y a en fait deux couples de solutions : une valeur de  $x_0$  avec une valeur de  $\delta$  positive, une autre valeur de  $x_0$  avec une valeur de  $\delta$  négative, il faut *a priori* choisir le signe de  $\delta$ .

En supposant  $\delta \geq 1$ , nous suivons la même marche que ci-dessus, avec :

$$\Gamma(\delta_y + 1) = \frac{1}{n} \Sigma |x_i - x_0| / SD^{\delta_e},$$

ou :

$$\delta_y = SL/(\text{Log} SD - C).$$

Le sens de variation de  $x_0$  suivant le signe de  $(\delta_y - \delta_e)$  est le même que plus haut.

#### 8.4. — Estimation des paramètres, loi exponentielle généralisée tronquée avec troncature. Estimation par les moments.

Nous ne reviendrons pas sur les définitions du tronquage et de la troncature : cf. paragraphes 7.4., 1.4. et 2.5.3.

Nous nous plaçons dans le cas de lois en « J » ( $\delta \geq 1$ ) ou dans le cas de lois dont la fonction de densité admet une tangente infinie à l'origine ( $\frac{1}{2} < \delta < 1$ ).

Nous admettons que le paramètre de position  $x_0$ , borne inférieure de l'intervalle de définition de la variable, est connu par avance avec une précision suffisante. Le paramètre d'échelle  $s$  est positif, et nous posons :

$$S_1 = \Sigma (x_i - x_0), \quad S_2 = \Sigma (x_i - x_0)^2, \quad S_3 = \Sigma (x_i - x_0)^3,$$

sommes qui sont convenablement connues car les observations non faites ( $x_i$  peu différent de  $x_0$ ) et donc non utilisées interviennent pour peu de chose dans ces totaux : nous nous servons de toutes les observations disponibles sans faire de troncature à un seuil supérieur à celui qui est imposé par la sensibilité de l'appareil de mesure.

Nous écrivons les équations :

$$R_2 = S_2/S_1 = s \frac{\Gamma(2\delta + 1)}{\Gamma(\delta + 1)},$$

$$R_3 = S_3/S_2 = s \frac{\Gamma(3\delta + 1)}{\Gamma(2\delta + 1)},$$

d'où en éliminant  $s$  :

$$\Gamma(3\delta + 1) \Gamma(\delta + 1) = [\Gamma(2\delta + 1)]^2,$$

qui peut se mettre sous la forme :

$$\frac{2}{\sqrt{3}} \left( \frac{16}{27} \right)^\delta = \frac{\Gamma(\delta + 1/3) \cdot \Gamma(\delta + 2/3)}{[\Gamma(\delta + 1/2)]^2}.$$

et permet de déterminer  $\delta$  par approximations successives.

Le paramètre  $s$  est ensuite calculé par :

$$s = R_2 \frac{\Gamma(\delta + 1)}{\Gamma(2\delta + 1)}.$$

le nombre « théorique » d'observations est  $M$  donné par :

$$M = S_1/\Gamma(\delta + 1),$$

et si l'on appelle  $A$  le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, le paramètre  $F_0$  est calculé par :

$$F_0 = \frac{A - M}{A}.$$

### 8.5. — Détermination des paramètres par le maximum de vraisemblance. Loi exponentielle généralisée tronquée avec troncature.

La borne  $x_0$  est connue par avance, le seuil de troncature  $x_h$  est choisi (cf. paragraphe 7.4.), le paramètre d'échelle est supposé positif, ainsi que le paramètre de forme.

On applique la méthode aux  $n$  observations conservées ( $x_i \geq x_h$ ) dont la probabilité totale est :

$$(1 - F_0) e^{-\left(\frac{x_h - x_0}{s}\right)^{\frac{1}{\delta}}}.$$

Pour une valeur  $x_i$  de la variable, la fonction de densité s'écrit :

$$(1 - F_0) \frac{1}{s \delta} \left( \frac{x_i - x_0}{s} \right)^{\frac{1}{\delta} - 1} e^{-\left(\frac{x_i - x_0}{s}\right)^{\frac{1}{\delta}}} / (1 - F_0) e^{-\left(\frac{x_h - x_0}{s}\right)^{\frac{1}{\delta}}}.$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$-n \operatorname{Log} \delta - \frac{n}{\delta} \operatorname{Log} s + \left( \frac{1}{\delta} - 1 \right) \Sigma \operatorname{Log} (x_i - x_0) - s^{-\frac{1}{\delta}} \Sigma (x_i - x_0)^{\frac{1}{\delta}} + s^{-\frac{1}{\delta}} n (x_h - x_0)^{\frac{1}{\delta}}.$$

En dérivant par rapport à  $s$  et  $\delta$  et en annulant les dérivées, après avoir posé :

$$SD = \frac{1}{n} \Sigma (x_i - x_0)^{1/\delta},$$

$$SM = \frac{1}{n} \Sigma (x_i - x_0)^{1/\delta} \operatorname{Log} (x_i - x_0),$$

$$SL = \frac{1}{n} \Sigma \operatorname{Log} (x_i - x_0),$$

$$HD = (x_h - x_0)^{1/\delta},$$

$$HL = \operatorname{Log} (x_h - x_0),$$

il vient :

$$s = (SD - HD),$$

$$\delta = \frac{SM - HL \cdot HD}{SD - HD} - SL.$$

Cette dernière équation contient  $\delta$  comme seule inconnue. On pourrait la résoudre en choisissant une valeur de  $\delta$  pour calculer le second membre, d'où une valeur de  $\delta$  au premier membre, portée dans le second et ainsi de suite. La convergence est sûre mais lente. Le programme que nous avons écrit pour ordinateur inverse la fonction.

En appelant  $A$  le nombre total d'observations possibles, compte non tenu du tronquage ni de la troncature, le paramètre de tronquage  $F_0$  se calcule par :

$$F_0 = 1 - \frac{n/A}{e^{-\left(\frac{x_h - x_0}{s}\right)^{1/\delta}}}.$$

### 8.6. — Comportement asymptotique de la distribution exponentielle généralisée.

Si la fréquence  $F$  tend vers 1, et en posant  $T = \frac{1}{1-F}$ ,  $T$  étant le temps de récurrence ou période de retour, la variable réduite  $u$  tend vers :  $(\operatorname{Log} T)^\delta$  si  $\delta > 0$  et  $s > 0$ , tend vers :  $T^{-\delta}$  si  $\delta < 0$  et  $s > 0$ .

Si  $s$  est négatif et si la fréquence  $F$  tend vers zéro, en posant alors  $T = \frac{1}{F}$ , la variable réduite  $u$  tend vers :  $(\operatorname{Log} T)^\delta$  si  $\delta > 0$ , tend vers :  $T^{-\delta}$  si  $\delta < 0$ .

### 8.7. — Calculs à l'ordinateur.

Les paramètres  $x_0$ ,  $s$  et  $\delta$  étant connus, il est facile d'établir les sous-programmes de calcul d'une probabilité correspondant à une valeur donnée de la variable en utilisant la fonction de bibliothèque EXP (Z) ou d'inversion de la fonction en utilisant la fonction de bibliothèque ALOG (Z).

## IX. — DISTRIBUTION BÊTA INCOMPLÈTE

### 9.1. — Expressions mathématiques.

#### 9.1.1. — Fonction de répartition.

Nous l'écrivons :  $F(x) = \frac{1}{B(p, q)} \int_0^u u^{p-1} (1-u)^{q-1} du$  avec  $u = \frac{x-x_0}{s}$ .

$F(x)$  est la fréquence au non dépassement ; lorsque  $u$  varie de zéro à un,  $F(x)$  croît de zéro à un.  $B(p-q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ ,  $s$  est le paramètre d'échelle qui peut se mettre sous la forme  $s = x_1 - x_0$ ,  $x_1 > x_0$ . La variable  $x$  admet comme borne inférieure de son intervalle de définition (pour  $u = 0$ ) le paramètre de position  $x_0$  et comme borne supérieure (pour  $u = 1$ ) le paramètre  $x_1$ . Ces trois paramètres  $s$ ,  $x_0$  et  $x_1$  ont les mêmes dimensions que  $x$ , mais on doit n'en utiliser que deux conjointement.

$p$  et  $q$  sont les paramètres de forme, positifs, différents de zéro.

9.1.2. — La fonction de densité est  $f(u) = \frac{1}{B(p, q)} u^{p-1} (1-u)^{q-1}$  dont la dérivée première s'écrit :

$$\frac{1}{B(p-q)} u^{p-2} (1-u)^{q-2} [(p-1) - u(p+q-2)], \text{ de racine } u = \frac{p-1}{p+q-2},$$

et la dérivée seconde :

$$\frac{1}{B(p, q)} u^{p-3} (1-u)^{q-3} [(p-1)(p-2) - 2u(p-1)(p+q-3) + u^2(p+q-2)(p+q-3)],$$

de racines  $u = \frac{1}{p+q-2} \left[ (p-1) \pm \sqrt{\frac{(p-1)(q-1)}{p+q-3}} \right]$

d'où les différentes formes possibles de la représentation graphique de la fonction de densité, qui est symétrique par rapport à  $u = \frac{1}{2}$  si  $p = q$ .

Nous allons passer en revue ces différentes formes en supposant  $q > p$ . Si  $q$  est plus petit que  $p$ , il suffit de considérer la figure symétrique par rapport à  $u = \frac{1}{2}$  : en faisant le changement de variable,  $v = 1 - u$ ;  $p$  joue le rôle de  $q$  et inversement.

$0 < p < 1$ ,  $p \leq q < 1$  : densités de probabilité infinies à  $u = 0$  et  $u = 1$ .

Contre-mode ( $f(u)$  minimal) à  $u = (p-1)/(p+q-2)$ . Pas de point d'inflexion. Modes observés à  $u = 0$ ,  $u = 1$ .

$q = 1$  : densité de probabilité infinie à  $u = 0$ , égale à  $p$  à  $u = 1$ . Pas de mode réel ni de point d'inflexion. Mode observé à  $u = 0$ .

$1 < q < 2 - p$  : densité de probabilité infinie à  $u = 0$ , nulle à  $u = 1$  avec tangente infinie. Pas de mode réel, un point d'inflexion à :

$$u = \frac{1}{p+q-2} \left[ (p-1) - \sqrt{\frac{(p-1)(q-1)}{p+q-3}} \right]. \text{ Mode observé à } u = 0.$$

$q = 2 - p$  : densité de probabilité infinie à  $u = 0$ , nulle à  $u = 1$  avec tangente infinie. Pas de mode réel, un point d'inflexion à  $u = 1 - p/2$ . Mode observé à  $u = 0$ .

$2 - p < q < 2$  : densité de probabilité infinie à  $u = 0$ , nulle à  $u = 1$  avec tangente infinie. Pas de mode réel, un point d'inflexion à :

$$u = \frac{1}{p + q - 2} \left[ (p - 1) + \sqrt{\frac{(p - 1)(q - 1)}{p + q - 3}} \right]. \text{ Mode observé à } u = 0.$$

$q = 2$  : densité de probabilité infinie à  $u = 0$ , nulle à  $u = 1$  avec tangente  $-p(p + 1)$ . Pas de mode réel ni de point d'inflexion. Mode observé à  $u = 0$ .

$2 < q$  : densité de probabilité infinie à  $u = 0$ , nulle à  $u = 1$  avec tangente nulle. Pas de mode réel ni de point d'inflexion. Mode observé à  $u = 0$ .

$p = 1, q = 1$  : densité de probabilité constante, égale à 1 : c'est la distribution continue, uniforme.

$1 < q < 2$  : densité de probabilité égale à  $+q$  pour  $u = 0$  et nulle pour  $u = 1$  avec tangente infinie. Pas de mode réel ni de point d'inflexion. Mode observé à  $u = 0$ .

$q = 2$  : densité de probabilité égale à 2 pour  $u = 0$  et nulle pour  $u = 1$  avec tangente  $-2$ . La fonction de densité est la droite  $f(u) = 2(1 - u)$ . Mode observé à  $u = 0$ .

$2 < q$  : densité de probabilité égale à  $+q$  pour  $u = 0$  et nulle pour  $u = 1$  avec tangente nulle. Pas de mode réel ni de point d'inflexion. Mode observé à  $u = 0$ .

$1 < p < 2, p \leq q < 2$  : densité de probabilité nulle à  $u = 0$  avec tangente infinie, de même à  $u = 1$ . Mode réel à  $u = \frac{p - 1}{p + q - 2}$ . Pas de point d'inflexion.

$q = 2$  : densité de probabilité nulle à  $u = 0$  avec tangente infinie, nulle à  $u = 1$  avec tangente  $-p(p + 1)$ . Mode à  $u = \frac{p - 1}{p}$ , pas de point d'inflexion.

$2 < q$  densité de probabilité nulle à  $u = 0$  avec tangente infinie, nulle à  $u = 1$  avec tangente nulle. Mode à  $u = \frac{p - 1}{p + q - 2}$ , un point d'inflexion à

$$\frac{1}{p + q - 2} \left[ (p - 1) + \sqrt{\frac{(p - 1)(q - 1)}{p + q - 3}} \right].$$

$p = 2, q = 2$  : densité de probabilité à  $u = 0$  avec tangente  $+6$ , nulle à  $u = 1$  avec tangente  $-6$ . Mode à  $u = \frac{1}{2}$ . Pas de point d'inflexion.

$2 < q$  : densité de probabilité nulle à  $u = 0$  avec tangente  $+q(q + 1)$ , nulle à  $u = 1$  avec tangente nulle. Mode à  $u = \frac{1}{q}$ . Point d'inflexion à  $u = \frac{2}{q}$ .

$2 < p, p \leq q$  : densité de probabilité nulle à  $u = 0$  et  $u = 1$  avec tangente nulle. Mode à  $u = (p - 1)/(p + q - 2)$ . Points d'inflexion à

$$u = \frac{1}{p + q - 2} \left[ (p - 1) \pm \sqrt{\frac{(p - 1)(q - 1)}{p + q - 3}} \right]. \text{ Courbe en cloche.}$$

Nous montrons en figure 16 les différentes formes des représentations graphiques de la fonction de répartition et de la fonction de densité d'après les valeurs de  $u$ .

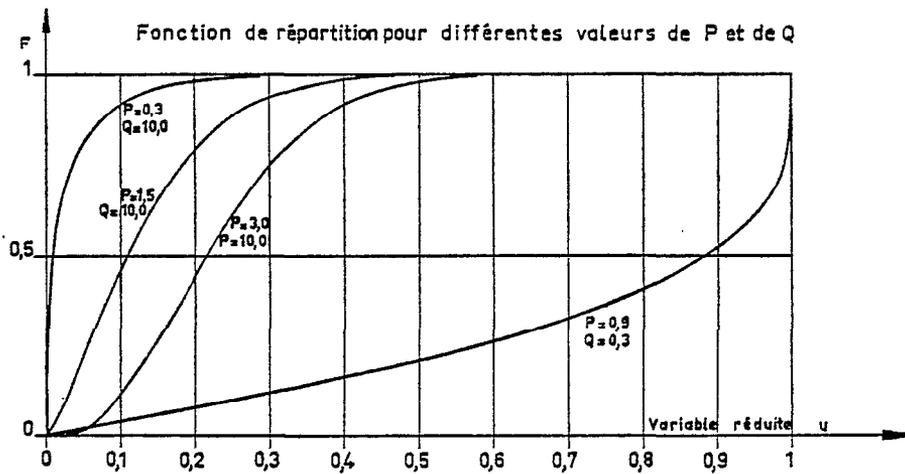
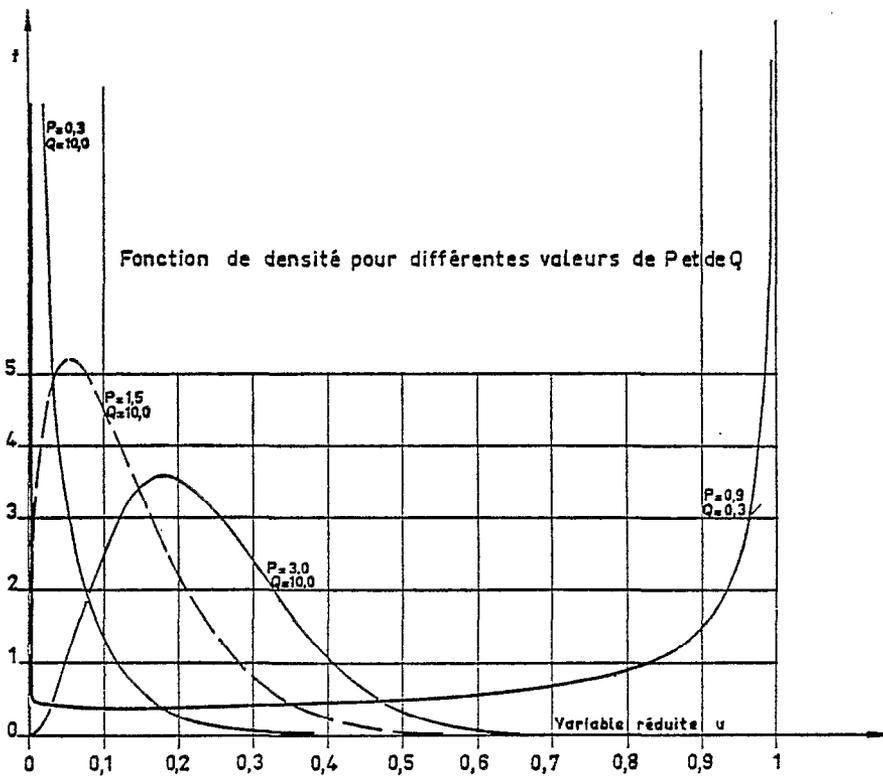


FIG. 16.

9.1.3. — Moments, cumulants, valeurs centrales d'une distribution bêta incomplète, parfaitement connue.

Le moment non centré d'ordre  $r$  de la variable réduite est :

$$m_r = \frac{\Gamma(p+r)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+r)} = \frac{p(p+1)\dots(p+r-1)}{(p+q)(p+q+1)\dots(p+q+r-1)}.$$

d'où :

le moment d'ordre 1, moyenne,  $p/(p+q)$  ;

le moment centré d'ordre 2, variance,  $k_2 = pq/(p+q)^2(p+q+1)$  ;

le cumulant d'ordre 3,  $k_3 = -2pq(p-q)/(p+q)^3(p+q+1)(p+q+2)$  ;

le cumulant d'ordre 4,  $k_4 = \frac{6pq[(p+q)^2(p+q+1) - pq[5(p+q)+6]]}{(p+q)^4(p+q+1)^2(p+q+2)(p+q+3)}$ .

Dans le cas de lois Bêta symétriques ( $p=q$ ), les moments centrés d'ordre impair sont nuls, et les moments centrés d'ordre pair ( $2r$ ) sont :

$$\mu_{2r} = \frac{1}{2^{2r}} \frac{(2r-1)!}{(r-1)!} \frac{(p+1)\dots(p+r-1)}{(2p+1)\dots(2p+2r-1)}.$$

Le coefficient d'asymétrie est :  $\gamma_1 = \frac{2(q-p)}{p+q+2} \sqrt{\frac{p+q+1}{p+q}}$ .

L'asymétrie est du signe de  $(q-p)$ .

Le coefficient d'aplatissement est :

$$\gamma_2 = 6 \frac{(p+q)^2(p+q+1) - pq[5(p+q)+6]}{pq(p+q+2)(p+q+3)}.$$

Dans le cas de distributions Bêta symétriques, le premier est évidemment nul, et le second vaut  $-\frac{6}{2p+3}$  : ces distributions sont plus aplaties que la distribution normale, vers laquelle elles tendent lorsque  $p$  tend vers l'infini.

Si  $q$  seul tend vers l'infini,  $p$  restant fini, la variable  $u$  tend à suivre une loi gamma incomplète de paramètres  $\gamma = p$ ,  $s = 1/q$  et  $x_0 = 0$ .

Le mode est :

$$x_0 + s \frac{p-1}{p+q-2} = x_1 - s \frac{q-1}{p+q-2} = \frac{x_1(p-1) + x_0(q-1)}{p+q-2}$$

n'existant que pour  $p$  et  $q > 1$ .

La moyenne est :

$$x_0 + s \frac{p}{p+q} = x_1 - s \frac{q}{p+q} = \frac{x_1 p + x_0 q}{p+q}.$$

La moyenne harmonique est, pour la variable réduite, de  $\frac{p-1}{p+q-1}$  qui n'a de sens que pour  $p > 1$ . L'inverse de la moyenne harmonique est  $\frac{p+q-1}{p-1}$  dont la variance  $\frac{q(p+q-1)}{(p-1)^2(p-2)}$  n'a de sens que pour  $p > 2$ .

La moyenne logarithmique (logarithme naturel de la moyenne géométrique) est, pour la variable réduite, de  $\psi(p) - \psi(p+q)$  et sa variance est  $\psi'(p) - \psi'(p+q)$ .

### 9.2. — Estimation des paramètres par les moments (échantillon de taille $n$ ).

En utilisant les sommes des quatre premières puissances des valeurs observées ( $x_1 \dots x_i \dots x_r$ ) et en posant :

$$S_1 = \frac{1}{n} \Sigma x_i, \quad S_2 = \frac{1}{n} \Sigma x_i^2, \quad S_3 = \frac{1}{n} \Sigma x_i^3, \quad S_4 = \frac{1}{n} \Sigma x_i^4,$$

les bornes  $x_0$  et  $x_1$  ( $x_1 > x_0$ ) sont les racines de l'équation du second degré en  $Y$  :

$$\begin{aligned} & Y^2 [-2 (S_4 - S_3 S_1) (S_2 - S_1^2) - 6 (S_3 S_1 - S_2^2) (S_2 - S_1^2) + 3 (S_3 - S_2 S_1)^2], \\ & + Y [6 (S_4 S_1 - S_3 S_2) (S_2 - S_1^2) - 3 (S_3 S_1 - S_2^2) (S_3 - S_2 S_1) - (S_4 - S_3 S_1) (S_3 - S_2 S_1)], \\ & + (S_4 - S_3 S_1) (S_3 S_1 - S_2^2) - 3 (S_4 S_2 - S_3^2) (S_2 - S_1^2) = 0. \end{aligned}$$

$p$  et  $q$  sont donnés par :

$$\begin{aligned} p + q &= \frac{S_1 (x_0 + x_1) - x_0 x_1 - S_2}{S_2 - S_1^2}, \\ p &= (p + q) \frac{S_1 - x_0}{x_1 - x_0}, \quad q = (p + q) \frac{x_1 - S_1}{x_1 - x_0}. \end{aligned}$$

9.2.1. — Si une borne est choisie *a priori*, l'autre est déterminée par l'équation :

$$\begin{aligned} & x_0 x_1 [2 S_1 (S_2 - S_1^2) - (S_3 - S_2 S_1)], \\ & + (x_0 + x_1) [S_1 (S_3 - S_2 S_1) - 2 S_2 (S_2 - S_1^2)], \\ & + 2 S_3 (S_2 - S_1^2) - S_2 (S_3 - S_2 S_1) = 0, \end{aligned}$$

$p$  et  $q$  étant déterminés comme ci-dessus.

Si l'échelle  $s = x_1 - x_0$  a été choisie *a priori*, il suffit de porter dans l'équation ci-dessus  $x_1 = s + x_0$  pour avoir une équation du second degré en  $x_0$  dont les deux racines sont les bornes cherchées.

9.2.2. — Si les deux bornes ont été choisies *a priori* (ou une borne et le paramètre d'échelle),  $p$  et  $q$  sont déterminés comme ci-dessus.

### 9.3. — Détermination des paramètres par le maximum de vraisemblance (échantillon de taille $n$ ).

Pour une valeur  $x_i$  de la variable, la fonction de densité s'écrit :

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{x_i - x_0}{s}\right)^{p-1} \left(1 - \frac{x_i - x_0}{s}\right)^{q-1}.$$

Le logarithme naturel du maximum de vraisemblance s'écrit :

$$\begin{aligned} \text{Log } \Gamma(p+q) - \text{Log } \Gamma(p) - \text{Log } \Gamma(q) + \frac{p-1}{n} \Sigma \text{Log } (x_i - x_0) + \frac{q-1}{n} \Sigma \text{Log } (x_1 - x_i) \\ - (p+q-1) \text{Log } (x_1 - x_0), \end{aligned}$$

en dérivant par rapport à  $x_0$ ,  $x_1$ ,  $p$  et  $q$  et en annulant les dérivées :

$$\text{(dérivée par rapport à } x_0) \quad \frac{p-1}{n} \sum \frac{1}{x_i - x_0} = \frac{p+q-1}{x_1 - x_0},$$

$$\text{(dérivée par rapport à } x_1) \quad \frac{q-1}{n} \sum \frac{1}{x_1 - x_i} = \frac{p+q-1}{x_1 - x_0},$$

$$\psi(p+q) - \psi(p) + \frac{1}{n} \sum \text{Log}(x_i - x_0) - \text{Log}(x_1 - x_0) = 0,$$

$$\psi(p+q) - \psi(q) + \frac{1}{n} \sum \text{Log}(x_1 - x_i) - \text{Log}(x_1 - x_0) = 0,$$

La première équation est inutilisable si  $p \leq 1$  et la seconde si  $q \leq 1$ . Nous proposons de les remplacer respectivement par :

si  $p \leq 1$   $S_1 - x_0 = (x_1 - x_0) \frac{p}{p+q}$  en posant  $S_1 = \frac{1}{n} \sum x_i$ .

si  $q \leq 1$   $x_1 - S_1 = (x_1 - x_0) \frac{q}{p+q}$ ,

et si  $p$  et  $q$  sont tous les deux  $\leq 1$ , en posant  $S_2 = \frac{1}{n} \sum x_i^2$

nous proposons de remplacer les deux premières équations par :

$$p+q = \frac{S_1(x_0 + x_1) - x_0 x_1 - S_2}{S_2 - S_1^2},$$

et l'une des deux équations précédentes ou :

$$p-q = (p+q) \frac{2S_1 - (x_0 + x_1)}{x_1 - x_0}.$$

#### 9.4. — Calculs à l'ordinateur.

9.4.1. — Calcul de la probabilité correspondant à une valeur donnée de la variable.

Il s'agit d'utiliser un développement limité permettant de calculer :

$$F(x) = \frac{1}{B(p, q)} \int_0^u u^{p-1} (1-u)^{q-1} du \text{ avec } u = \frac{x-x_0}{s}.$$

Nous faisons un changement de variable en posant :

$$d = \left[ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \right]^{\frac{1}{p+q-1}} \text{ et } v = d u,$$

d'où :

$$F(x) = \int_0^v v^{p-1} (d-v)^{q-1} dv,$$

en intégrant par parties :

$$F(x) = \frac{v^p}{p} (d-v)^{q-1} \sum_1^{\infty} U_i.$$

Les termes de la série  $U$  se déduisent les uns des autres, à partir du premier  $U_1 = 1$  par la formule de récurrence :

$$U_{i+1} = \frac{q-i}{p+i} \frac{v}{d-v} U_i$$

cette série est convergente si  $v < \frac{d}{2}$  ( $u < \frac{1}{2}$ ) et même alternée lorsqu'on arrive au terme dont le rang est supérieur à la partie entière de  $q$ .

Si  $u$  est supérieur à  $1/2$ , il suffit de faire le changement de variable  $u' = 1 - u$  et d'intervertir  $p$  et  $q$  dans le programme.

Si le terme multiplicateur de la série est inférieur à une valeur choisie d'après la précision demandée sur la probabilité, nous ne faisons pas le calcul de la série mais prenons  $F(x) = 0$  ou  $F(x) = 1$  suivant que  $x$  est inférieur ou supérieur à la moyenne. Le calcul de la valeur maximale de  $v$  permettant d'assimiler  $F(x)$  à zéro se fait en considérant :  $\int_0^v v^{p-1} (d-v)^{q-1} dv$  comme la surface d'un triangle de base  $v$  et hauteur  $v^{p-1} (d-v)^{q-1}$ .

9.4.2. — Calcul de la valeur de la variable correspondant à une probabilité donnée.

L'inversion de la fonction Bêta incomplète se fait facilement par la méthode générale « du pas » en partant de la valeur moyenne de la variable réduite :

$$v = \frac{p}{p+q}$$

### 9.5. — Relations de la distribution bêta incomplète avec d'autres distributions.

Avec la loi binomiale : la somme des termes du rang  $a$  au rang  $n$  d'une loi binomiale :

$$\sum_{i=a}^n \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

est égale à :

$$\frac{1}{B(a, n+1-a)} \int_0^p u^{a-1} (1-u)^{n-a} du.$$

Avec la distribution de Student : la probabilité (au non dépassement) définie par la distribution de Student est égale à la probabilité au dépassement, calculée d'après une loi Bêta incomplète de paramètres :

$$p = \frac{v}{2} \quad q = \frac{1}{2}, \text{ de bornes zéro et un, et de variate } u = \frac{v}{v+t^2}.$$

Avec la distribution  $F$  de Fisher-Snédecor :  $F$  étant le rapport de variances  $s_1^2/s_2^2$  (lois normales) de degrés de libertés respectifs  $v_1$  et  $v_2$ , la probabilité (au non-dépassement) définie par la valeur de  $F$  est égale à la probabilité, au dépassement, calculée d'après une loi Bêta incomplète de paramètres  $p = \frac{v_2}{2}$   $q = \frac{v_1}{2}$  de bornes zéro et un, et de variate  $\frac{v_2}{v_2 + v_1 F}$ .

## X. — CHOIX D'UNE LOI

### 10.1. — Prolifération des lois de probabilité.

Pendant longtemps le principal critère du choix d'une loi de probabilité par un utilisateur a été la possibilité d'emploi de cette loi : existence de tables de la fonction de distribution, facilité de calcul des paramètres d'ajustement à la main ou à la machine de bureau. L'usage d'un ordinateur supprime le recours aux tables et permet l'estimation de paramètres par les méthodes les plus efficaces qui souvent ne peuvent être utilisées pour le calcul à la main.

L'habitude est prise de n'utiliser en hydrologie que des lois de probabilité absolument continues : on ne peut nier le caractère « discret » des grandeurs que nous observons, et même discret avec peu de chiffres significatifs étant donné la précision de nos appareils de mesure. Si l'on continue à préférer les distributions absolument continues, le nombre de lois proposées et utilisées ne cessera d'augmenter. Même si l'on s'en tient aux conditions les plus dures pour la définition mathématique de la fonction de répartition : lorsque la variable parcourt en croissant tout son intervalle de définition, la relation doit être biunivoque entre la variable et la fonction qui elle-même doit être monotone et croissante de 0 à 1 ; cette fonction admet une dérivée continue toujours  $\geq 0$ , appelée fonction de densité, laquelle ne doit avoir au plus qu'un seul maximum ou un seul minimum entre les bornes non comprises de l'intervalle de définition.

### 10.2. — Choix d'une loi parmi celles que nous avons étudiées.

Le choix d'une loi de probabilité risque de devenir de plus en plus difficile à mesure que le nombre des fonctions proposées augmentera. Tant que l'on s'en tiendra à des lois ayant au plus deux paramètres de forme (la loi de Halphen du type A en a trois) — c'est-à-dire qu'il suffit d'au plus des quatre premiers moments pour déterminer tous les paramètres — un graphique du genre de la figure 17 pourra être un guide.

10.2.1. — Ce graphique montre, pour les quelques lois que nous avons étudiées, les variations de l'un par rapport à l'autre des coefficients  $\gamma_1$  d'asymétrie et  $\gamma_2$  d'aplatissement, qui sont tous les deux indépendants des paramètres de position et d'échelle et dont la valeur ne dépend — quelle que soit la distribution choisie — que des paramètres de forme :

$$\gamma_1 = K_3/K_2^{3/2}$$

$$\gamma_2 = K_4/K_2^2$$

les cumulants étant calculés pour un échantillon de taille  $n$  et en posant :

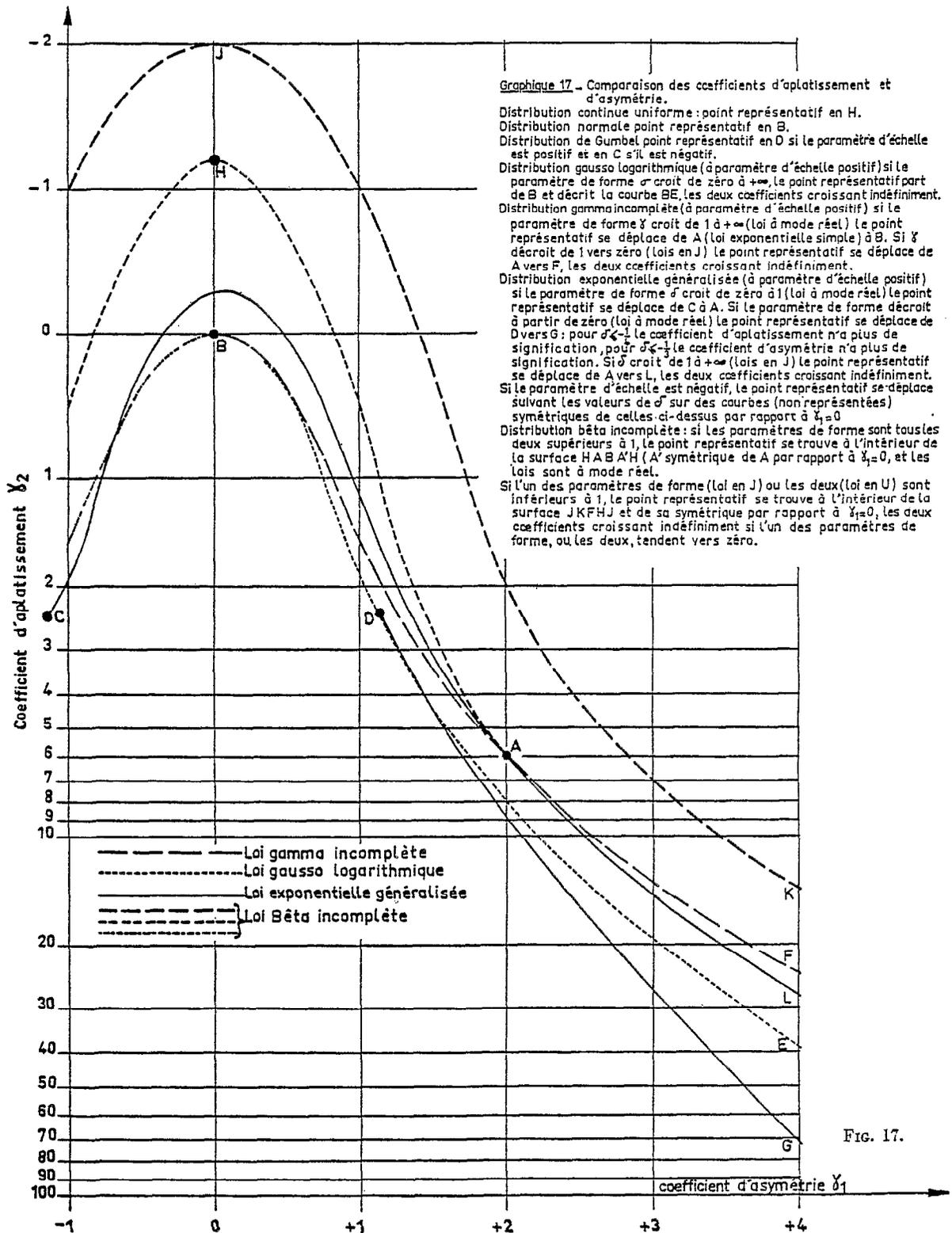
$$S_1 = \Sigma x_i, S_2 = \Sigma x_i^2, S_3 = \Sigma x_i^3, S_4 = \Sigma x_i^4,$$

par :

$$K_2 = \frac{1}{n-1} \left( S_2 - \frac{S_1^2}{n} \right).$$

$$K_3 = \frac{1}{(n-1)(n-2)} \left( n S_3 - 3 S_2 S_1 + \frac{2}{n} S_1^3 \right)$$

$$K_4 = \frac{1}{(n-1)(n-2)(n-3)} \left[ (n^2 + n) S_4 - 4(n+1) S_3 S_1 - 3(n-1) S_2^2 + 12 S_2 S_1^2 - \frac{6}{n} S_1^4 \right].$$



Rappelons que le coefficient d'asymétrie est positif si la distribution est étalée sur la droite, négatif si elle est étalée sur la gauche, que le coefficient d'aplatissement est positif si la distribution est moins aplatie que la distribution normale, négatif si elle est plus aplatie.

10.2.2. — Pour bien faire, il faudrait compléter le graphique de la figure 17 par des bandes représentant différents intervalles de confiance des coefficients  $\gamma_1$  et  $\gamma_2$  pour différentes tailles d'échantillon.

La formulation mathématique de ces intervalles de confiance semble quasiment impossible à établir, mais en utilisant une méthode de Monte-Carlo nous n'aurions pas de difficultés à les tracer (calculs par ordinateurs).

10.2.3. — On peut aussi penser, pour le choix de la loi la mieux représentative, à l'utilisation de tests d'ajustement en comparant les valeurs trouvées de ces tests pour les différentes lois essayées. Nous renvoyons au paragraphe 2.3. où le sujet des tests d'ajustement a été traité.

## XI. — INVERSION DE FONCTIONS

11.1. — Soit  $F(x)$  une fonction quelconque de la variable  $x$ , on cherche la valeur  $x_s$  de la variable pour laquelle  $F(x_s) = Y$ ,  $Y$  étant une valeur de la fonction, donnée à l'avance.

On se place dans le cas où la relation  $Y = F(x)$  est biunivoque, c'est-à-dire que l'équation  $F(x) - Y = 0$  n'admet qu'une racine réelle.

On rencontre souvent ce cas, non seulement dans le problème de l'inversion des fonctions de distribution, mais aussi dans le calcul des paramètres par la méthode du maximum de vraisemblance, soit que l'on puisse en isoler un (distribution de Gumbel, distribution gaussio-logarithmique, distribution exponentielle généralisée tronquée) soit que l'on reste en présence d'un système d'équations (distribution gamma incomplète tronquée ou non tronquée, distribution exponentielle généralisée non tronquée).

Plusieurs méthodes sont possibles.

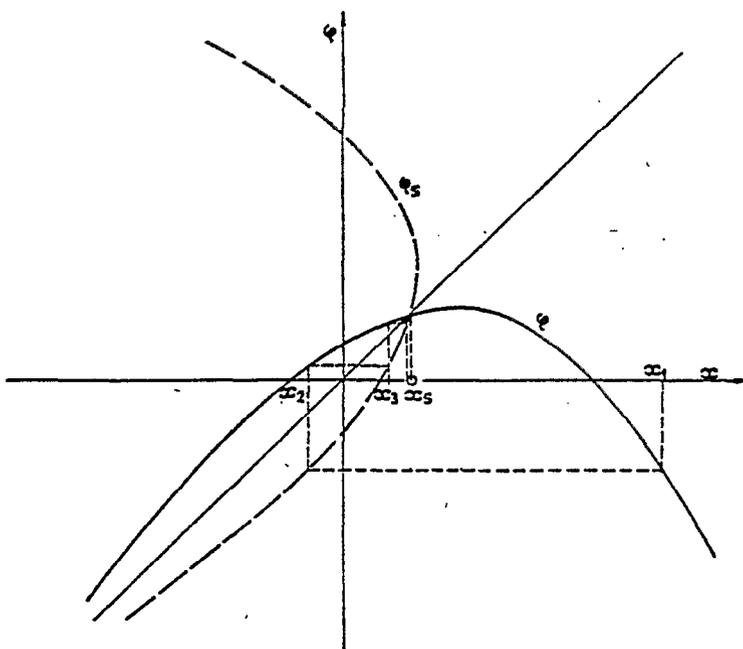


FIG. 18.

11.2. — Ecrivons  $x = F(x) - Y + x$  c'est-à-dire  $x = \varphi(x)$ .

Il est tentant de partir en itération directe : choisir une valeur  $x_1$ , calculer  $\varphi(x_1) = x_2$  ; calculer  $\varphi(x_2) = x_3 \dots$  jusqu'à trouver, si l'on est dans les conditions d'auto-convergence (cf. plus bas), une valeur  $x_n$  différant de la valeur  $x_{n-1}$  de l'erreur relative ou absolue que l'on a admise *a priori* : la solution est  $x_s = x_n$ .

Graphiquement (fig. 18), la marche suivie est simple si l'on considère  $\varphi$  et sa symétrique  $\varphi_s$  par rapport à la première bissectrice. On en déduit la condition d'auto-convergence de l'itéra-

tion directe quel que soit  $x$  : en tout point de  $\varphi(x)$  la tangente doit être comprise entre  $-1$  et  $+1$  :

$$-1 < \varphi'(x) < +1 \quad \text{c'est-à-dire} \quad -2 < F'(x) < 0 \quad \text{quelque soit } x.$$

Cette méthode est d'une écriture très simple en FORTRAN, mais ne s'applique pas souvent. Notamment, elle n'est pas utilisable pour l'inversion des fonctions de répartition pour lesquelles  $F'(x) = f(x) > 0$ . De plus, la convergence semble assez lente lorsque  $\varphi'(x_0)$  est voisin de  $-1$  ou de  $+1$ .

### 11.3. — Méthode la plus générale.

Soit la fonction  $F(x)$ , que nous savons calculer en fonction de  $x$  pour toute valeur comprise dans l'intervalle de définition de la variable. Nous cherchons  $x_s$  tel que  $F(x_s) = Y$ ,  $Y$  étant une valeur donnée à l'avance.

Nous choisissons une valeur initiale  $x_1$ , par exemple la moyenne, la médiane, le mode ou autre, conduisant à un calcul facile de  $F(x_1)$  que nous comparons à  $Y$ . Si la différence entre  $Y$  et  $F(x_1)$  est supérieure à ce que nous avons admis en erreur absolue ou relative, nous allons essayer une nouvelle valeur  $x_2$  pour calculer  $F(x_2)$ .

Trois cas sont à considérer :

#### Cas A.

L'intervalle de définition n'est pas borné : on a avantage à choisir — si possible — la moyenne comme valeur initiale  $x_1$ , à prendre comme nouvelle valeur  $x_2 = x_1 \pm \sqrt{\text{variance}}$ , le signe étant celui de  $(Y - F(x_1))$  et à prendre 2 comme valeur du pas  $p$  qui servira aux itérations ultérieures.

#### Cas B.

L'intervalle de définition est borné d'un seul côté, inférieurement par exemple, par la valeur  $x_0$ . Prendre comme nouvelle valeur  $x_2 = (x_1 - x_0)p + x_0$ , le pas  $p$  valant 0,5 si  $(Y - F(x_1))$  est négatif et 2 si  $(Y - F(x_1))$  est positif.

#### Cas C.

L'intervalle de définition est borné inférieurement par une valeur  $x_0$  et supérieurement par une valeur  $x_t$ . Prendre comme nouvelle valeur  $x_2 = (x_1 - x_0)p + x_0$  ou  $x_2 = x_t - (x_t - x_1)p$  (avec  $p = 0,5$ ) suivant que  $(Y - F(x_1))$  est négatif ou positif.

On explore l'intervalle de définition de la variable en comparant  $Y$  aux valeurs de  $F(x_n)$  calculées pour les valeurs successives de  $x_n$  : d'après les formules de récurrence.

$$\text{Cas A : } x_n = (x_{n-1} - x_1)p + x_1 ;$$

$$\text{Cas B : } x_n = (x_{n-1} - x_0)p + x_0 ;$$

$$\text{Cas C : } x_n = (x_{n-1} - x_0)p + x_0 ;$$

$$\text{ou : } x_n = (x_{n-1} - x_t)p + x_t.$$

en répétant l'itération tant que  $(Y - F(x_n))$  reste de même signe : dès le premier changement de signe de cette quantité, on sait que la solution  $x_s$  se trouve comprise entre  $x_n$  et  $x_{n-1}$ .

On peut alors travailler encore en itération, suivant le processus : faire  $x_{n+1} = (x_n + x_{n-1}) / 2$ , comparer  $Y$  et  $F(x_{n+1})$ , choisir celui des  $x_n$  ou  $x_{n-1}$  pour lequel  $(Y - F(x_i))$  est de signe contraire à  $(Y - F(x_{n+1}))$ , faire  $x_{n+2} = (x_n + x_i) / 2$ , comparer  $Y$  et  $F(x_{n+2})$ , etc.

On peut aussi, et l'écriture sera plus rapide en FORTRAN, poser  $u_i = x_i - x_1$  (ou  $x_i - x_0$  ou  $x_i - x_t$ ) et réaliser l'itération  $u_{n+1} = u_n(1 + sp')$ .  $s$  est un signe  $+$  ou  $-$  définissant le sens de variation de  $u$  et déterminé par  $(Y - F(u_n))$ , et  $p'$  est un pas que l'on divise par 2 à chaque changement de signe de  $(Y - F(u_n))$ .

11.3.1. — Nous allons développer cette méthode générale d'inversion sur la fonction Bêta incomplète. On appellera le programme d'inversion **FUNCTION VBETA** (FX, P, Q, S, XO) (Tableau de la figure 19) ; VBÊTA est la valeur demandée de la variable pour la valeur donnée FX de la probabilité au non dépassement, P et Q sont les paramètres de forme, S est le paramètre d'échelle et XO le paramètre de position (borne inférieure de l'intervalle de définition de la variable).

Le problème consiste à trouver la valeur de la variable réduite UI pour la fréquence FX dans une loi Bêta de paramètres P, Q, 1 et 0. On en déduit la valeur cherchée VBETA (ligne 22, instruction n° 20).

SR et XR sont les paramètres d'échelle et de position de la variable réduite.

T est un indicateur de signe pouvant prendre la valeur + 1 ou la valeur - 1, initialisé à + 1.

```

FUNCTION VBETA(FX,P,Q,S,XQ)
SR=1.
XR=0
T=1.
IND=1
UI=P/(P+Q)
3 FI=FBETA(UI,P,Q,SR,XR)
IF(IFIX((FX-FI)*1.E7)*T)5,20,6
5 GO TO (11,12,18),IND
6 GO TO (10,14,15),IND
10 T=-1.
11 IND=2
12 UA=UI
UI=UI-(UI-.5+.5*T)/2.
IF((UI-1.E-7)*(1.-UI-1.E-7))20,20,3
14 IND=3
15 UB=UI
16 UI=(UA+UB)/2.
IF((UI-UA)*(UI-UB))3,20,20
18 UA=UI
GO TO 16
20 VBETA=UI*S+XQ
RETURN
END

```

FIG. 19.

IND est un indicateur d'aiguillage pouvant prendre les valeurs 1, 2 ou 3, et initialisé à 1.

Nous calculons une première valeur de UI (ligne 6), valeur moyenne de la variable réduite dans la loi Bêta (P, Q), puis (instruction n° 3, ligne 7) la fréquence FI correspondant à UI que nous comparons (ligne 8) à la fréquence FX.

Si FI ne diffère de FX que de ( $\pm .0000001$ ), nous admettons que UI est la valeur cherchée pour la variable réduite et nous allons à l'instruction n° 20.

Sinon, l'indicateur d'aiguillage prend la valeur 2, T garde la valeur + 1 si FI est trop grand ou prend la valeur - 1 si FI est trop petit, et ne changera plus de signe.

Nous gardons en réserve, sous l'appellation UA, la valeur précédemment calculée de UI (instruction n° 12), et calculons une nouvelle valeur de UI qui devient UA/2 si T = + 1 ou + .5 + UA/2 si T = - 1 (ligne 14), et nous retournons à l'instruction n° 3 pour calculer la fréquence FI correspondant à la nouvelle valeur de UI, comparer FX et FI à la ligne 8. Tant que le signe de FX - FI ne change pas, la valeur de l'indicateur ne change pas, nous allons à l'instruction n° 12, calculons une nouvelle valeur de UI et retournons à l'instruction n° 3 sauf si (ligne 15) UI ne diffère de zéro ou de un que de .0000001 ; dans ces cas, la valeur de UI est prise comme solution.

Si le signe de FX - FI change, l'indicateur d'aiguillage prend la valeur 3 (instruction n° 14) et nous savons que la solution se trouve entre UA et UI, dernière valeur calculée que nous gardons en réserve sous l'appellation UB (instruction n° 15), calculons  $UI = (UA + UB)/2$ , retournons à l'instruction n° 3 et le travail se poursuit en itération décrite à l'avant-dernier alinéa du paragraphe

précédent jusqu'à ce que (ligne 8) FI ne diffère de FX que de .0000001 ou que (ligne 19) l'ordinateur ne puisse plus distinguer de différence entre UI et une des valeurs précédentes de UI étant donné le nombre de chiffres significatifs qu'il conserve. UI est alors pris comme solution.

11.4. — Dans certains cas on peut utiliser des méthodes de calcul plus rapides à l'ordinateur : nécessitant environ trois fois moins d'itérations que la méthode ci-dessus, mais chaque itération étant un peu plus longue.

11.4.1. — Méthode des deux cordes. — Il faut d'abord explorer la fonction  $F(u)$ ,  $U$  étant la variable réduite, de façon à trouver trois abscisses (cf figure 20) du même côté par rapport au mode : UI, UJ et UK correspondant aux probabilités respectives FI, FJ et FK telles qu'une seule de ces probabilités — FJ — se trouve de l'autre côté de la probabilité FX à inverser par rapport à la probabilité modale (parce que, dans le cas de la figure 20, la pente de la tangente décroît lorsque  $U$  croît).

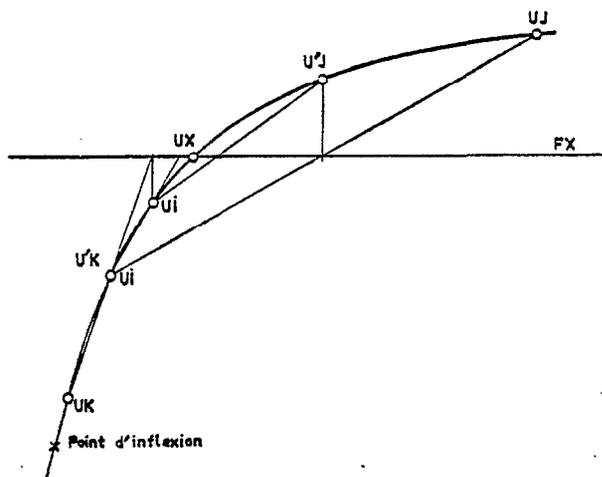


FIG. 20.

On trace les cordes UK UI et UI UJ qui coupent FX aux abscisses U'I et U'J, on calcule les probabilités F'I et F'J correspondant à ces deux abscisses, probabilités que l'on compare à FX. Si l'une de ces probabilités est égale à FX à l'erreur près que l'on s'est fixée à l'avance, le problème est résolu.

Si non, en appelant U'K la valeur primitive UI et F'K la valeur primitive FI, on recommence l'itération avec U'I U'J et U'K.

La convergence est rapide, mais il faut faire attention à la valeur de l'erreur absolue exigée sur FX car les déterminations de U'I et U'J se font par règle de trois. Si l'on fait les calculs en simple précision et qu'on inverse une fonction de probabilité, il faut accepter une erreur absolue de .000001 sur FX, sans cela on risque de déterminer U'I ou U'J du mauvais côté de UX.

Méthode de la corde unique. — C'est la méthode précédente avec suppression des points UK. Le point UI est fixe et le point U'J se rapproche de UX à chaque itération. La convergence est lente, d'autant plus lente que FI est différent de FX. Toutes choses égales par ailleurs, la précision ne peut être supérieure à celle obtenue par la méthode des deux cordes.

11.4.2. — Si l'on peut calculer rapidement la pente de la tangente à la fonction de probabilité, sans risquer de dépassement de la capacité de l'ordinateur pour un des termes du calcul de la pente, une méthode à convergence très rapide est celle de la tangente et corde.

Il faut d'abord explorer la fonction  $F(u)$ ,  $u$  étant la variable réduite, de façon à trouver deux abscisses (cf figure 21) du même côté par rapport au mode : UI et UJ dont les probabilités correspondantes FI et FJ encadrent la probabilité FX à inverser.

On trace la tangente en UI (UI étant l'abscisse la plus proche du mode) et la corde UI UJ qui coupent FX aux abscisses U'I et U'J. On calcule les probabilités F'I et F'J correspondantes que l'on compare à FX. Si l'une de ces probabilités est égale à FX à l'erreur près que l'on s'est fixée à l'avance, le problème est résolu. Sinon on recommence l'itération avec U'I et U'J.

La convergence est rapide, mais la précision est limitée de la même façon et pour les mêmes raisons qu'avec la méthode des deux cordes. Nous avons utilisé cette méthode pour l'inversion de la loi normale avec une erreur absolue de .0000001 sur FX car on obtient facilement et rapidement les valeurs des pentes et des probabilités en double précision.

Méthode de la tangente simple : c'est la méthode précédente avec suppression du point UJ. Convergence rapide, méthode à utiliser si le calcul de la probabilité est lent — même précision que ci-dessus.

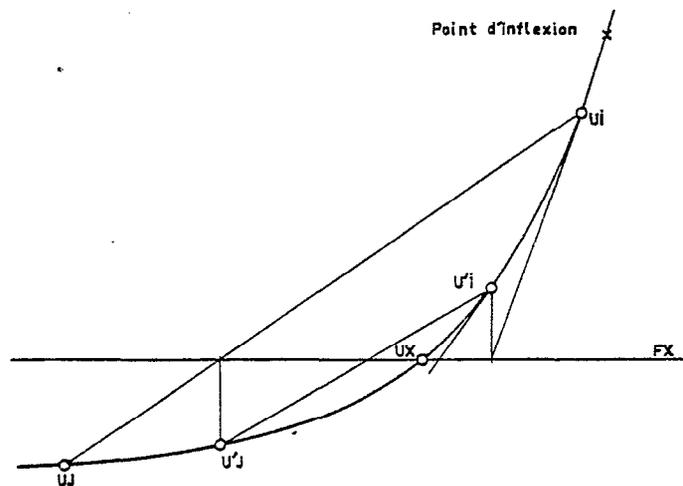


FIG. 21.

## XII. — PROGRAMMES FORTRAN

Nous donnons en annexe différents sous-programmes écrits en FORTRAN IV pour le niveau G des machines I.B.M. 360-50 et 75. Ces sous-programmes peuvent être utilisés directement au niveau H, mais auraient besoin de petites retouches pour correspondre au niveau E (machine I.B.M. 360-40).

### 12.1. — Distribution normale.

12.1.1. — FUNCTION FNORM Fonction de répartition de la loi normale.

12.1.2. — FUNCTION VNORM Inversion de la distribution normale.

12.1.3. — SUBROUTINE PNORM Calcul des paramètres d'une distribution normale, non tronquée, adaptée à un échantillon de taille connue, sans troncature.

### 12.2. — Distribution de Gumbel.

12.2.1. — FUNCTION FGUMB Fonction de répartition de la loi de Gumbel.

12.2.2. — FUNCTION VGUMB Inversion de la distribution de Gumbel.

12.2.3. — SUBROUTINE PGUMB Calcul des paramètres d'une distribution de Gumbel, non tronquée, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance (cf paragraphe 5.4.).

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, et la taille N de l'échantillon.

Le sous-programme calcule les valeurs du paramètre de position  $X_0 = x_0$ , du paramètre d'échelle  $S = s$  et de la moyenne de l'échantillon  $\bar{X}M$ .

Nous définissons, au début du sous-programme :

- un indicateur d'aiguillage IND valeur initiale 1 ;
- un indicateur de signe D valeur initiale + 1 ;
- un pas d'exploration BS valeur initiale .5

Nous calculons les moments d'ordre 1 (moyenne  $\bar{X}M$ ), 2 et 3 et le cumulatif d'ordre 3 dont le signe est E, et une valeur approchée  $s_1$  de S par la méthode des moments.

Puis (instructions n° 11), nous calculons :

$$Z = \sum e^{-\frac{x_i}{s_1}} \quad \text{et} \quad V = s_1 + \frac{\sum x_i e^{-\frac{x_i}{s_1}}}{z}$$

nous comparons  $V$  à  $XM$  (pour faire ensuite varier  $s$  suivant le schéma de la figure 22) en valeur relative et en signe par le calcul de  $U = (V - XM) / (ABS(V) + ABS(XM))$ , ce dénominateur étant choisi pour éviter une division par zéro.

L'instruction suivante IF(IFIX ...) permet de sortir de la boucle si l'erreur relative entre  $V$  et  $XM$  est inférieure à environ .000002 en valeur absolue, sinon nous remontons à l'instruction n° 6 et nous définissons une nouvelle valeur de l'indicateur d'aiguillage  $IND = 2$ , de l'indicateur de signe  $D$  qui prend le signe de  $U$ , et la valeur initiale du pas d'exploration  $AS$  (.5 si  $D * E$  est + 1., et 2. si  $D * E$  est - 1.).

Nous modifions la valeur de  $S$  en la multipliant par  $AS$  : c'est-à-dire si  $U$  est positif (parce que la valeur approchée  $s_1$ , était supérieure à la valeur de solution) et si  $S$  est positif, la nouvelle valeur adoptée est la valeur  $s_1$  divisée par 2 ; si  $S$  est négatif la nouvelle valeur adoptée est la valeur  $s_1$  multipliée par 2 (instruction n° 9).

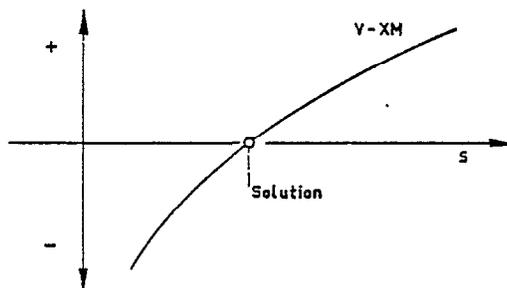


FIG. 22.

Nous calculons à nouveau  $Z$  et  $V$  et comparons  $V$  à  $XM$ . Si le signe de  $U$  n'a pas changé depuis la fois précédente, nous retournons à l'instruction 9 pour continuer à faire varier  $S$  dans le même sens, et la valeur de l'indicateur d'aiguillage ne change pas.

Si le signe de  $U$  a changé, nous allons à l'instruction n° 17 ; l'indicateur d'aiguillage prend la valeur  $IND = 3$ ,  $D$  change de signe, le pas  $BS$  est divisé par 2 et  $S$  croît ou décroît suivant que  $U$  est négatif ou positif (instruction n° 19).

Nous calculons à nouveau  $Z$  et  $V$  et comparons  $V$  à  $XM$ . Si le signe de  $U$  n'a pas changé depuis la fois précédente, nous allons à l'instruction n° 19 sans modifier la valeur du pas  $BS$ , et  $S$  continue à varier dans le même sens.

Si le signe de  $U$  a changé, nous allons à l'instruction n° 17 ;  $D$  change de signe, le pas  $BS$  est divisé par 2 et  $S$  varie en sens inverse du passage précédent mais avec une « accélération » diminuée de moitié (par exemple si  $S$  s'augmentait (ou se diminuait) de  $1/16$  au passage précédent,  $S$  se diminuera (ou s'augmentera) de  $1/32$  à ce nouveau passage).

Nous calculons à nouveau  $Z$  et  $V$  et comparons  $V$  à  $XM...$ , etc.

Le travail en boucle avec  $IND = 3$  continue jusqu'à ce que l'erreur relative en valeur absolue de  $V$  par rapport à  $XM$  soit inférieure à .000002 ou jusqu'à ce que le pas  $BS$  soit inférieur à .0000001 : c'est-à-dire dernière valeur de  $BS$  :  $1/2 \cdot 2^{-19}$  car l'ordinateur ne fait pas (en simple précision) la distinction entre 1 et  $1 + 1/2$  (à la puissance 20 ou 21).

Dans un cas comme dans l'autre, nous admettons que les dernières valeurs calculées de  $S$  et de  $Z$  le sont avec suffisamment de précision et nous calculons  $XO$  juste avant RETURN.

### 12.3. — Distribution gaussio-logarithmique.

12.3.1. — FUNCTION FLOGN Fonction de distribution de la loi gaussio-logarithmique.

12.3.2. — FUNCTION VLOGN Inversion de la distribution gaussio-logarithmique.

12.3.3. — SUBROUTINE PLOGN Calcul des paramètres d'une distribution gaussio-logarithmique, non tronquée, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance (cf. paragraphe 6.3.).

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, la taille N de l'échantillon et la valeur XS limite inférieure acceptée à priori pour XO (par exemple XS = 0 pour une étude de débits) inférieure à la valeur observée minimale.

Le sous-programme calcule les valeurs du paramètre de position XO = x<sub>0</sub> (borne inférieure de l'intervalle de définition de la variable, de valeur égale ou supérieure à XS), du paramètre d'échelle S = s positif, du paramètre de forme SI = σ positif, et de la moyenne de l'échantillon XM.

Nous définissons au début du sous-programme un indicateur d'aiguillage IND valeur initiale 1, un pas d'exploration AS valeur initiale 1. ; nous calculons la moyenne XM et déterminons la plus petite valeur observée XI.

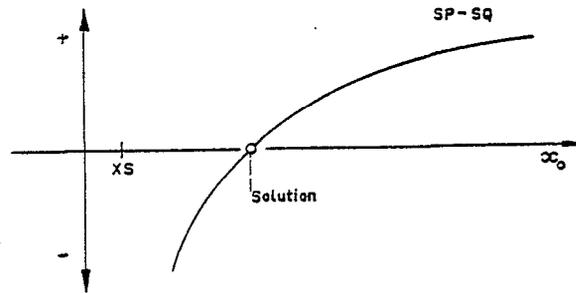


FIG. 23.

Nous choisissons une valeur initiale pour XO, inférieure à XI :

$x_{01} = XI - 2. * (XM - XI)$  et nous calculons

$$SQ = \frac{1}{n} \sum \text{Log}^2(x_i - x_{01}) - \frac{1}{n^2} (\sum \text{Log}(x_i - x_{01}))^2,$$

et

$$SP = \frac{1}{n} \sum \text{Log}(x_i - x_{01}) - \sum \frac{\text{Log}(x_i - x_{01})}{x_i - x_{01}} \bigg/ \sum \frac{1}{x_i - x_{01}}.$$

Nous comparons, dans l'instruction n° 19, SP et SQ en valeur relative :  $\frac{SP - SQ}{SQ}$  (SQ positif, ne peut être nul que si toutes les valeurs de  $x_i$  sont identiques entre elles) pour faire ensuite varier  $x_0$  suivant le schéma de la figure 23. L'instruction n° 19, IF (IFIX ...) correspondante permet de sortir de la boucle si l'erreur relative entre SP et SQ est inférieure à .00001 en valeur absolue, sinon :

- Si SP — SQ est positif, l'indicateur d'aiguillage ne change pas de valeur, nous retournons à l'instruction n° 12 et nous choisissons une nouvelle valeur de XO inférieure à la précédente  $XI - AS * (XM - XI)$ , AS étant à chaque passage multiplié par 2. Nous calculons SQ et SP et nous comparons ces deux valeurs comme plus haut. Tant que SP — SQ est positif, nous retournons à l'instruction n° 12, sauf si XO à force de diminuer (instruction n° 20) devient inférieur au seuil limite fixé XS. Dans ce cas, nous remontons à l'instruction n° 13, l'indicateur d'aiguillage prend la valeur IND = 0, XO prend la valeur XS et nous calculons les deux paramètres SI et S restant à déterminer, juste avant RETURN.
- Si SP — SQ est négatif, l'indicateur d'aiguillage prend la valeur IND = 3 et nous choisissons une nouvelle valeur de XO supérieure à la précédente  $XI + .5 * (XM - XI)$ . Nous calculons SQ et SP et nous comparons ces deux valeurs comme plus haut. Tant que SP — SQ

reste négatif nous retournons à l'instruction n° 12 et choisissons un XO supérieur au précédent  $XI - AS * (XM - XI)$ , AS étant à chaque passage divisé par 2, calculons SQ et SP ... etc ... sauf si à force de diminuer la quantité  $AS * (XM - XI)$  devient trop petite pour se retrancher effectivement de XI étant donné les possibilités de l'ordinateur en simple précision : l'instruction, qui précède immédiatement l'instruction n° 13, arrête le calcul. Il semble d'ailleurs impossible d'arriver à une quantité  $AS * (XM - XI)$  aussi petite avec les valeurs usuelles que nous rencontrons pour le paramètre de forme.

Lorsque à un moment de cette première exploration  $SP - SQ$  a changé de signe, nous donnons à l'indicateur la valeur  $IND = 4$  si  $SP - SQ$  devient positif ou  $IND = 2$  si  $SP - SQ$  devient négatif, nous remontons à l'instruction n° 14, et modifions XO précédent, dans le sens voulu, de la quantité  $AS * (XM - XI)$ , AS étant un pas moitié du dernier pas utilisé. Nous calculons SQ et SP, les comparons dans l'instruction n° 19 qui nous renvoie à l'instruction n° 14 sans changer la valeur de IND si  $SP - SQ$  n'a pas changé de signe, et refaisons une nouvelle boucle en modifiant XO sans modifier le signe mais en divisant par deux la valeur du pas d'exploration ; ou qui nous renvoie à l'instruction n° 14 si  $SP - SQ$  a changé de signe (IND passe alors de la valeur 2 à la valeur 4 ou inversement) et nous refaisons une nouvelle boucle en modifiant XO par un pas divisé par 2 et changé de signe.

Le travail en boucle se continue jusqu'à ce que l'erreur relative entre SP et SQ soit inférieure à .00001 en valeur absolue ou jusqu'à ce que le pas d'exploration devienne trop petit pour modifier effectivement la valeur de XO étant donné les possibilités de l'ordinateur en simple précision.

Dans les deux cas nous admettons que les dernières valeurs calculées de XO, SQ et SL le sont avec suffisamment de précision, et nous calculons les paramètres SI et S juste avant RETURN.

12.3.4. — SUBROUTINE PLOGNP calcul des paramètres inconnus d'une distribution gaussio-logarithmique, non tronquée, adaptée à un échantillon de taille connue, sans troncature et dont on connaît la borne limite inférieure (paramètre de position), par la méthode du maximum de vraisemblance.

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, la taille N de l'échantillon et la valeur XO du paramètre de position, valeur inférieure à la valeur observée minimale.

Le sous-programme calcule les valeurs du paramètre d'échelle  $S = s$ , positif, du paramètre de forme  $SI = \sigma$ , positif, et de la moyenne de l'échantillon XM.

12.3.5. — FUNCTION FLOGNT Fonction de distribution de la loi gaussio-logarithmique tronquée.

12.3.6. — FUNCTION VLOGNT Inversion de la distribution gaussio-logarithmique tronquée.

12.3.7. — SUBROUTINE PLOGNT Calcul des paramètres inconnus d'une distribution gaussio-logarithmique tronquée adaptée à un échantillon avec troncature (cf paragraphe 6.5.). On suppose connue la valeur du paramètre de position, et les paramètres de forme, d'échelle et de troncature seront déterminés, les deux premiers par la méthode du maximum de vraisemblance.

Le programme principal doit fournir la totalité des valeurs X de la variable (y compris les valeurs inférieures au seuil de troncature) sous forme d'un vecteur, la taille N de l'échantillon, la valeur du paramètre de position XO, la valeur XH du seuil de troncature (qui pourrait être pris à  $(XO + \varepsilon)$  valeur de la plus petite des observations différente de XO).

Le sous-programme calcule les valeurs du paramètre de forme  $SI = \sigma$  positif, du paramètre d'échelle  $S = s$ , positif, et du paramètre de troncature  $FO = F_0$  compris entre 0 et 1.

Nous calculons d'abord les valeurs moyennes de  $\text{Log}(X(J) - XO)$  et de  $(\text{Log}(X(J) - XO))^2$  pour les NH observations supérieures ou égales à XH, ainsi que la proportion  $FH = NH/N$  de ces observations par rapport à la taille de l'échantillon.

Nous définissons un indicateur d'aiguillage JND, valeur initiale 1, un pas d'exploration de valeur initiale  $AS = 1/2$ . et une valeur initiale du paramètre de forme  $SI = 2$ . avant de nous rendre à l'instruction 13 où nous calculons une nouvelle valeur du paramètre de forme  $SI * pas$ ; ensuite à l'instruction 14 nous calculons  $SS = \text{Log}(S)$  d'après la valeur choisie de  $SI$  et les valeurs connues (provenant indirectement du programme principal)  $SM, SL, UL$  — d'où une valeur  $S$  du paramètre d'échelle.

Nous calculons alors la quantité  $XL$  que nous comparons à  $SL$  dans l'instruction IF (IFIX...) qui permet de sortir de la boucle si l'erreur relative entre  $XL$  et  $SL$  est inférieure à .00001 en valeur absolue.

Si  $XL$  est inférieur à  $SL$ , il faut diminuer la valeur de  $SI$ . Ni l'indicateur d'aiguillage ni le pas ne changent de valeurs et nous retournons à l'instruction 13 et refaisons les calculs ci-dessus jusqu'à ce que  $XL$  devienne supérieur à  $SL$ .

Si  $XL$  est supérieur à  $SL$ , il faut augmenter la valeur de  $SI$ . L'indicateur d'aiguillage prend la valeur 3, le pas prend la valeur 2. et nous retournons à l'instruction 12 où nous mettons en mémoire (valeur  $SA$ ) la valeur de  $SI$  avant de passer à l'instruction 13, puis refaisons les calculs ci-dessus jusqu'à ce que  $XL$  devienne inférieur à  $SL$ .

La valeur du paramètre de forme se trouve alors comprise entre la dernière valeur  $SB$  utilisée comme approximation de ce paramètre et la valeur  $SA$  utilisée au tour précédent. L'indicateur d'aiguillage prend la valeur 2 et nous utilisons comme nouvelle valeur approchée de  $SI$  la quantité  $(SA + SB)/2$  pour retourner à l'instruction 14.

La boucle de calcul continue en prenant pour valeur de  $SI$  la valeur moyenne entre la dernière valeur utilisée et celle des deux valeurs précédentes pour laquelle le signe de  $(XL - SL)$  est inversé, jusqu'à ce que l'on en sorte — soit par l'instruction IF (IFIX...) soit lorsque les possibilités de la machine en simple précision ne lui permettent plus de distinguer entre deux valeurs successives de  $SI$ .

Le paramètre  $FO$  est alors calculé juste avant l'instruction RETURN.

#### 12.4. — Distribution gamma incomplète.

12.4.1. — FUNCTION FGAMA Fonction de distribution de la loi gamma incomplète, non tronquée.

12.4.2. — FUNCTION VGAMA Inversion de la distribution gamma incomplète, non tronquée.

12.4.3. — SUBROUTINE PGAMA Calcul des paramètres d'une distribution gamma incomplète, non tronquée, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance chaque fois que cela est possible, sinon en remplaçant l'équation provenant de la dérivée par rapport à  $x_0$  par l'équation contenant le moment de second ordre (cf. paragraphe 7.3.).

Le programme principal doit fournir les valeurs  $X$  de la variable sous forme d'un vecteur, la taille  $N$  de l'échantillon et la valeur  $XS$  limite acceptée a priori pour  $XO$  (par exemple  $XS = 0$ ,  $XS = -1$ . E 50) inférieure à la valeur minimale observée dans l'échantillon.

Le sous-programme calcule les valeurs du paramètre de position  $XO = x_0$  (borne inférieure de l'intervalle de définition de la variable, de valeur égale ou supérieure à  $XS$ ), du paramètre d'échelle  $S = s$ , positif, du paramètre de forme  $GC = \gamma$ , positif et de la moyenne  $XM$  de l'échantillon.

Nous commençons, dans le sous-programme, par calculer le moment d'ordre 1 (moyenne), le moment d'ordre 2 et nous déterminons la valeur minimale observée  $XI$ .

Puis le même système d'itérations va nous servir pour estimer des paramètres soit en utilisant les trois équations dérivées du maximum de vraisemblance, soit en remplaçant une de ces équations par celle du moment de second ordre. Les quantités que nous appelons  $\gamma_y$  et  $\gamma_z$  dans le

paragraphe 7.3. sont dénommées GC et GZ dans le sous-programme. La figure 24 montre les sens de variation de GZ-GC lorsque GZ est calculé par le maximum de vraisemblance et quand cette méthode convient (fig. 24-1), quand cette méthode ne convient pas (fig. 24-2) et lorsque GZ est calculé par le moment de second ordre (fig. 24-3), cas où l'on a toujours une solution quel que soit  $\gamma$ .

Nous définissons un premier indicateur JND de valeur initiale 1 qui correspond à la recherche de la solution dans le cas de la figure 24-1 et qui prendra la valeur 2 pour le cas de la figure 24-2.

Nous définissons aussi un indicateur d'aiguillage IND valeur initiale 1 et un pas d'exploration AS valeur initiale 1.

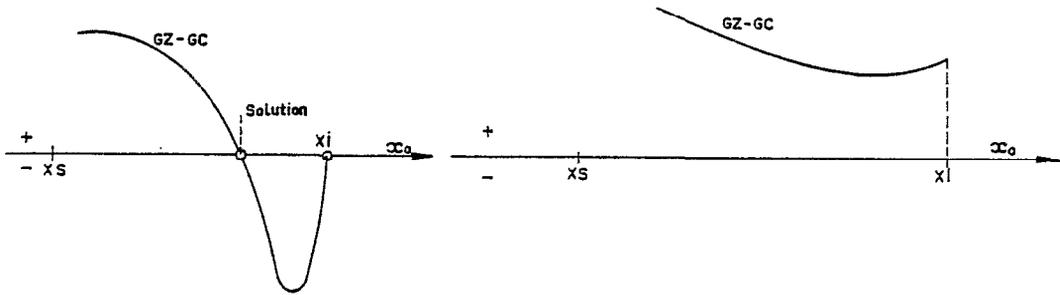


FIG. 24-1.

FIG. 24-2.

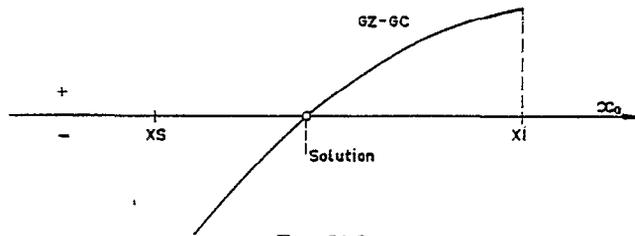


FIG. 24-3.

Nous choisissons une valeur initiale pour XO, inférieure à XI,  $x_{01} = XI - \frac{1}{2} * (XM - XI)$  et nous calculons GZ (instruction n° 16) et GC.

Nous comparons, dans l'instruction n° 22, GZ et GC en valeur relative  $\frac{GZ - GC}{GC}$  (GC, positif, ne peut pas être nul) pour faire ensuite varier XO dans le sens voulu. Cette instruction n° 22 par laquelle on repasse à chaque tour de boucle de calcul, permet de terminer en sortant de la boucle si l'erreur relative entre GZ et GC est inférieure à .0001 en valeur absolue sinon :

- Si  $GZ - GC$  est positif, l'indicateur d'aiguillage ne change pas de valeur, nous retournons à l'instruction n° 10 et nous choisissons une valeur de XO supérieure à la précédente  $XI - AS * (XM - XI)$ , AS étant à chaque passage divisé par 2. Nous calculons GZ et GC et nous comparons ces deux valeurs comme plus haut. Tant que  $GZ - GC$  reste positif nous retournons à l'instruction n° 10, calculons GZ et GC..., etc. sauf si, à force de diminuer, la quantité  $AS * (XM - XI)$  devient trop petite pour se retrancher effectivement de XI étant donné les possibilités de l'ordinateur en simple précision : nous admettons alors que nous sommes dans le cas de la figure 24-2 et que la solution correspond à la figure 24-3. Nous faisons alors  $JND = 2$  et retournons à l'instruction n° 8.
- Si  $GZ - GC$  est négatif, l'indicateur d'aiguillage prend la valeur  $IND = 3$ , et nous choisissons une nouvelle valeur de XO inférieure à la précédente  $XI - 2 * (XM - XI)$ . Nous calculons GZ et GC et nous comparons ces deux valeurs comme plus haut. Tant que  $GZ - GC$

reste négatif, nous retournons à l'instruction n° 12 et nous choisissons une nouvelle valeur de XO inférieure à la précédente  $XI - AS * (XM - XI)$ , AS étant multiplié par 2 à chaque passage. Nous calculons GZ et GC et nous comparons ces deux valeurs comme plus haut..., etc., sauf si XO à force de diminuer (instruction n° 23) devient inférieur au seuil limite fixé XS. Dans ce cas, nous remontons à l'instruction n° 11, l'indicateur d'aiguillage prend la valeur  $IND = 0$ , XO prend la valeur XS et nous calculons les deux paramètres restant à déterminer GC (instruction n° 18 sqq) et S, juste avant RETURN.

Lorsque à un moment de cette première exploration ( $JND = 1$ ,  $IND = 1$  puis 3)  $GZ - GC$  a changé de signe, est devenu positif, nous donnons à l'indicateur la valeur  $IND = 4$ , remontons à l'instruction n° 12 et modifions le XO précédent de la quantité  $AS * (XM - XI)$ , AS étant un pas moitié du dernier pas utilisé. Nous calculons GZ et GC, les comparons dans l'instruction n° 22 qui nous renvoie à l'instruction n° 12 ( $IND$  prenant la valeur 2 si  $GZ - GC$  a changé de signe). Nous allons faire des séries de boucles avec  $IND = 4$  ou  $IND = 2$ , la valeur de  $IND$  changeant à chaque changement de signe de  $GZ - GC$ , ces boucles se faisant en modifiant XO sans modifier le signe du pas si  $IND$  ne change pas de valeur, et en prenant un pas divisé par 2 à chaque retour.

Le travail en boucle se continue jusqu'à ce que l'erreur relative entre  $(GZ - GC)$  et GC soit inférieure à .0001 en valeur absolue, ou jusqu'à ce que le pas d'exploration devienne trop petit pour modifier effectivement la valeur de XO étant donné les possibilités de l'ordinateur en simple précision.

Dans les deux cas, nous admettons que les dernières valeurs calculées de XO et de GC le sont avec suffisamment de précision et nous calculons le paramètre S juste avant RETURN.

Dans le cas où JND a pris la valeur 2, la marche du calcul est semblable à celle décrite ci-dessus, avec comme différence importante l'inversion du signe de  $GZ - GC$  pour déterminer le sens de variation de XO.

12.4.4. — SUBROUTINE PGAMAP Calcul des paramètres inconnus d'une distribution gamma incomplète, non tronquée, adaptée à un échantillon de taille connue, sans troncature, et dont on connaît la borne limite inférieure (paramètre de position) — par la méthode du maximum de vraisemblance.

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, la taille N de l'échantillon et la valeur XO du paramètre de position, valeur inférieure à la valeur observée minimale.

Le sous-programme calcule les valeurs du paramètre d'échelle  $S = s$ , positif, du paramètre de forme  $GC = \gamma$ , positif, et de la moyenne de l'échantillon XM.

12.4.5. — FUNCTION FGAMAT Fonction de distribution de la loi gamma incomplète tronquée.

13.4.6. — FUNCTION VGAMAT Inversion de la distribution gamma incomplète tronquée.

12.4.7. — SUBROUTINE PGAMAT Calcul des paramètres d'une distribution gamma incomplète, tronquée, adaptée à un échantillon avec troncature, dont on connaît le paramètre de position.

Un tel sous-programme peut se déduire facilement du suivant : SUBROUTINE PGAMAJ. L'expérience a prouvé que ce sous-programme PGAMAT ne fonctionnait pas bien lorsque le paramètre de forme était supérieur à 1,5 ; nous avons cherché, sans y parvenir, une méthode de solution valable quel que soit  $\gamma$  du système d'équation du paragraphe 7.5.

12.4.8. — SUBROUTINE PGAMAJ calcul des paramètres d'une distribution gamma incomplète tronquée, adaptée à un échantillon avec troncature (cf. paragraphe 7.5.). On suppose connue la valeur du paramètre de position, et les paramètres de forme, d'échelle et de tronquage seront déterminés, les deux premiers par la méthode du maximum de vraisemblance.

Le sous-programme a été établi pour l'étude de pluviométries journalières, mais peut être facilement modifié pour tout autre objet : les seuls changements à faire se trouvant dans les calculs préliminaires.

Le programme principal doit fournir : la valeur du paramètre de position XO et du seuil de troncature XH (qui pourrait être pris à  $(XO + \varepsilon)$ , valeur de la plus petite des observations différente de XO).

Les observations sont transportées, avec INTEGER\* 2 IX, ID en COMMON IX (70,366), ID (70) — IX (M, J) étant la hauteur pluviométrique correspondant au jour J de l'année dont les deux derniers chiffres du millésime forment M — ID (M) étant un indicateur d'utilisation de l'année M, valant 1 si les observations sont complètes et sans commentaires, — 1 si les observations de l'année sont totalement manquantes ou si elles ne sont pas conservées pour le calcul, et 0 si les observations sont utilisées malgré les commentaires (observations incomplètes, non quotidiennes, douteuses...).

Le sous-programme calcule les valeurs du paramètre de forme  $GO = \gamma$ , positif, du paramètre d'échelle  $SO = s = 1/a$ , positif et du paramètre de tronquage  $FO = F_0$  compris entre zéro et un.

Nous commençons par calculer dans le sous-programme : les valeurs des sommes de toutes les observations différentes de XO, de leurs carrés et de leurs cubes : d'où des valeurs approchées  $AO = \frac{1}{s}$  et GO par la méthode des moments ; le nombre relatif PH des observations supérieures ou égales à XH ainsi que leur moyenne SX et la moyenne SL de leurs logarithmes naturels (le nombre relatif PH est le nombre des observations  $\geq XH$  divisé par le nombre total d'observations possibles, ici somme des 365 ou 366 jours des années utilisées).

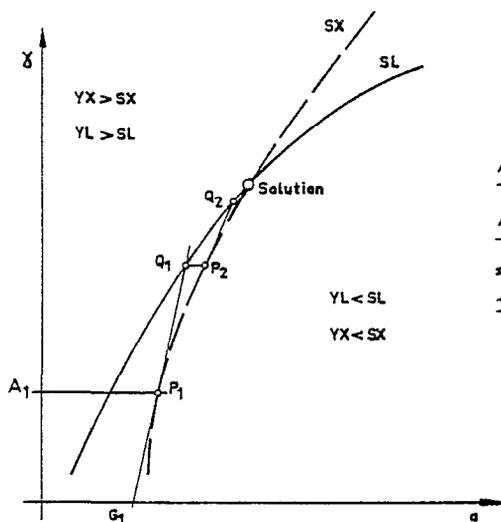


FIG. 25-1.

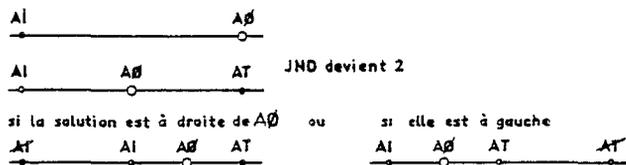


FIG. 25-2.

La méthode de travail du sous-programme est la suivante :

**Première boucle :** Nous explorons le plan  $a, \gamma$  (cf. figure 25-1) à  $\gamma$  constant jusqu'à trouver le point  $P_1$  tel que les valeurs de SX et de  $YX = \gamma/a + u^\gamma e^{-u}/[a \Gamma(\gamma) (1 - Pu)]$  soient les mêmes à une erreur relative que nous avons prise égale à .0001 (instruction IF (IFIX(ALOG(YX/SX))\* 1.E4).

Au tout premier départ, nous commençons avec les valeurs de  $\gamma$  et de  $a$  calculées par la méthode des moments, et avec l'indicateur à la valeur  $JND = 1$ . Si  $YX < SX$  il faut diminuer  $a$ ,

JND ne change pas de valeur et  $A_1 P_1 = AO$  est divisé par un pas de 2 tant que  $JND = 1$ . Si  $YX > SX$  il faut augmenter  $a$ , JND prend la valeur 3 et  $A_1 P_1$  est multiplié par un pas de 2 tant que  $JND = 3$ .

Dès que  $YX - SX$  a changé de signe, l'indicateur prend la valeur  $IND = 2$  et nous prenons comme valeur de  $a$  la moyenne entre la dernière valeur  $AO$  (conservée sous la dénomination  $AT$ ) et la valeur précédente  $AI$ .

La figure 25-2 montre le schéma de fonctionnement avec  $IND = 2$  tant que l'instruction  $IF(IFIX(ALOG \dots)$  n'a pas fait sortir de la boucle.

En sortant de la boucle, nous faisons  $JND = 4$  avant de rentrer dans la *seconde boucle* : nous explorons le plan  $a, \gamma$  (cf figure 25-1) sur la tangente en  $P_1$  à la courbe  $SX$  jusqu'à trouver le point  $Q_1$  tel que les valeurs de  $SL$  et de  $YL = -\text{Log } a + \psi(\gamma) / (1 - P_u) - SU / [\Gamma(\gamma) \times (1 - P_u)]$  soient les mêmes à une erreur relative que nous avons prise égale à  $.0001$  (instruction  $IF(IFIX(ALOG(YL/SL) * 1.E4)$ ).

Nous partons avec les valeurs de  $\gamma$  et de  $a$  provenant de première boucle, calculons la série :

$$SU = \sum \frac{(-1)^r}{r!} \frac{u_h^{\gamma+r}}{\gamma+r} \left( \text{Log } U_h - \frac{1}{\gamma+r} \right)$$

puis la pente de la tangente en  $P_1$  et la quantité  $YL$ .

Si  $YL > SL$ , il faut diminuer  $\gamma$ , JND prend la valeur 1 et tant que  $YL - SL$  ne change pas de signe, JND ne change pas de valeur et  $G_1 P_1$  est divisé avec un pas de 2.

Si  $YL < SL$  il faut augmenter  $\gamma$ , JND prend la valeur 3 et tant que  $YL - SL$  ne change pas de signe, JND ne change pas de valeur et  $G_1 P_1$  est multiplié par un pas de 2.

A chacune de ces itérations, non seulement  $\gamma$  change de valeur mais aussi  $a$  et il faut recalculer la série  $SU$ .

Dès que  $YL - SL$  change de signe, l'indicateur prend la valeur  $JND = 2$  et nous suivons un schéma très analogue à celui de la figure 25-2 tant que l'instruction  $IF(IFIX(ALOG \dots)$  n'a pas fait sortir de la boucle.

En sortant de la boucle, nous faisons  $JND = 4$  et retournons dans la première boucle : exploration à  $\gamma$  constant avec la valeur déterminée dans la seconde boucle ... Puis nous reprenons la seconde boucle ... Dès qu'on a fait un parcours dans une boucle sans changer la valeur de  $JND$  parce que  $YX$  se trouve à moins de  $0,01\%$  de  $SX$  et  $YL$  à moins de  $0,01\%$  de  $SL$  avec les mêmes valeurs de  $\gamma$  et de  $a$  qui sont les valeurs cherchées  $GO$  et  $AO = 1/SO$ , on passe à l'instruction 55 et on calcule le paramètre de tronquage  $FO$  juste avant  $RETURN$ .

## 12.5. — Distribution exponentielle généralisée.

12.5.1. — **FUNCTION FEXGE** fonction de distribution de la loi exponentielle généralisée, avec paramètres de forme et de position positifs ou négatifs, et non tronquée.

12.5.2. — **FUNCTION VEXGE** inversion de la distribution ci-dessus.

12.5.3. — **SUBROUTINE PEXGE** calcul des paramètres d'une distribution exponentielle généralisée, non tronquée, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance chaque fois que cela est possible, sinon en remplaçant l'équation provenant de la dérivée par rapport à  $x_0$  par l'équation contenant la moyenne logarithmique (cf paragraphe 8.3.).

Le programme principal doit fournir les valeurs  $X$  de la variable sous forme d'un vecteur, la taille  $N$  de l'échantillon, la valeur  $XS$  limite acceptée *a priori* pour  $XO$  (par exemple  $XS = 0$ ,  $XS = -1.E50$  ou  $XS = +1.E50$ , valeur qui doit être inférieure à la valeur minimale observée

dans l'échantillon si le paramètre d'échelle est positif ou supérieur à la valeur maximale observée dans l'échantillon si le paramètre d'échelle est négatif), le signe  $DS = + 1.$  ou  $- 1.$  du paramètre de forme et le signe  $SS = + 1.$  ou  $- 1.$  du paramètre d'échelle.

Le sous-programme calcule les valeurs du paramètre de position  $XO = x_0$  (borne inférieure de l'intervalle de définition de la variable si le paramètre d'échelle est positif,  $XO \geq XS$ , ou borne supérieure de l'intervalle de définition si le paramètre d'échelle est négatif,  $XO \leq XS$ ), du paramètre d'échelle  $S = s$ , avec son signe choisi à l'avance, du paramètre de forme  $DE = \delta$ , avec son signe choisi à l'avance, et de la moyenne  $XM$  de l'échantillon.

Nous commençons, dans le sous-programme, par calculer la moyenne  $XM$  et déterminer la valeur minimale (ou maximale) observée  $XI$ .

Puis le même système d'itérations va nous servir pour estimer les paramètres — soit en utilisant les trois équations dérivées du maximum de vraisemblance, soit en remplaçant une de ces équations par celle de la moyenne arithmétique ou celle de la moyenne logarithmique. Les quantités que nous appelons  $\delta_x$  et  $\delta_y$  dans le paragraphe 8.3 sont dénommées  $DE$  et  $DY$  dans le sous-programme. La figure 26 montre le sens de variation de  $(DY - DE)$  en fonction de  $XO$  ( $S$  étant positif, ce que nous allons supposer dans tout le raisonnement ci-dessous).

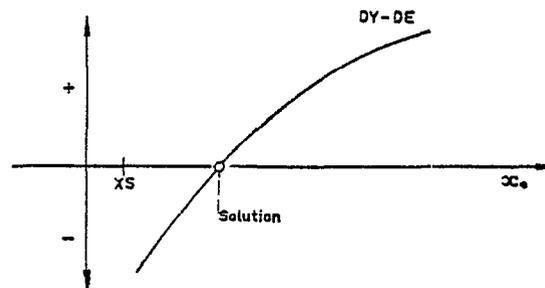


FIG. 26.

Nous définissons un premier indicateur JND de valeur initiale 1 qui correspond à la recherche de la solution par le maximum de vraisemblance et qui prendra la valeur 2 ou 3 lorsqu'on sera obligé de chercher la solution par l'intermédiaire de la moyenne arithmétique ou de la moyenne logarithmique.

Nous définissons aussi un indicateur d'aiguillage IND valeur initiale 1 et un pas d'exploration AS valeur initiale 1.

Nous choisissons une valeur initiale pour  $XO$ , inférieure à  $XI$ ,  $x_{01} = XI - \frac{1}{2} (XM - XI)$ , et nous calculons  $DE$ .

Pour cela, nous avons à résoudre l'équation implicite :

$$DE = \delta = \frac{\sum \text{Log} |x_i - x_0| \times |x_i - x_0|^{1/\delta}}{\sum |x_i - x_0|^{1/\delta}} - \frac{1}{n} \sum \text{Log} |x_i - x_0|.$$

C'est une inversion de fonction que l'on réalise facilement par la méthode de la valeur moyenne (cf paragraphe 11.3.) entre l'instruction n° 21 et l'instruction n° 29. A la première itération (premier essai avec  $JND = 1, 2$  ou  $3$ ) la valeur de départ est  $DE = DS = + 1.$  ou  $- 1.$ , aux itérations suivantes la valeur de départ est la valeur calculée pour  $DE$  à l'itération précédente.

Nous calculons  $DY$  (instructions nos 35, 36 ou 37) défini par une équation choisie d'après la valeur de  $DE$  (par l'intermédiaire de l'indicateur LND), et nous comparons  $DY$  et  $DE$  dans l'instruction n° 39 pour faire varier  $x_0$  dans le sens voulu. Cette instruction par laquelle on repasse à chaque tour de boucle de calcul permet de terminer en sortant de la boucle si l'erreur relative entre  $DY$  et  $DE$  est inférieure à .000001 en valeur absolue, sinon :

- Si  $DY - DE$  est négatif, l'indicateur d'aiguillage ne change pas de valeur, nous retournons à l'instruction n° 10 et nous choisissons une valeur de  $XO$  supérieure à la précédente :  $XI - AS^* (XM - XI)$ ,  $AS$  étant divisé par 2 à chaque passage. Nous calculons  $DY$  et  $DE$  et nous comparons ces deux valeurs comme plus haut. Tant que  $DY - DE$  reste négatif, nous retournons à l'instruction n° 10, calculons  $DY$  et  $DE$ , etc., sauf si, à force de diminuer, la quantité  $AS^* (XM - XI)$  devient trop petite pour se retrancher effectivement de  $XI$  étant donné les possibilités de l'ordinateur en simple précision : nous admettons alors que nous sommes dans le cas de  $\delta \geq 1$ . Nous faisons alors  $JND = 2$  et nous retournons à l'instruction n° 50.
- Si  $DY - DE$  est positif, l'indicateur d'aiguillage prend la valeur  $IND = 3$  et nous choisissons une nouvelle valeur de  $XO$  inférieure à la précédente  $XI - 2 \cdot (XM - XI)$ . Nous calculons  $DY$  et  $DE$  et nous comparons ces deux valeurs comme plus haut. Tant que  $DY - DE$  reste positif, nous retournons à l'instruction n° 12 et nous choisissons une nouvelle valeur de  $XO$  inférieure à la précédente :  $XI - AS^* (XM - XI)$ ,  $AS$  étant multiplié par 2 à chaque passage. Nous calculons  $DY$  et  $DE$  et nous comparons ces deux valeurs, etc., sauf si  $XO$  à force de diminuer (instruction n° 40) devient inférieur au seuil limite fixé  $XS$ . Dans ce cas nous remontons à l'instruction n° 14, l'indicateur d'aiguillage prend la valeur  $IND = 0$ ,  $XO$  prend la valeur  $XS$  et nous calculons  $DE$  puis le paramètre d'échelle  $S$  juste avant  $RETURN$ .

Lorsque à un moment de cette première exploration ( $JND = 1$ ,  $IND = 1$  puis 3)  $DY - DE$  a changé de signe, est devenu négatif, nous donnons à l'indicateur la valeur  $IND = 2$ , remontons à l'instruction n° 15 et modifions le  $XO$  précédent de la quantité  $AS^* (XM - XI)$ ,  $AS$  étant un pas moitié du dernier pas utilisé. Nous calculons  $DY$  et  $DE$ , les comparons dans l'instruction n° 23, qui nous renvoie à l'instruction n° 15 sans changer la valeur de l'indicateur si  $DY - DE$  n'a pas changé de signe, ou avec  $IND = 4$  si  $DY - DE$  a changé de signe : nous allons faire des séries de boucles avec  $IND = 2$  ou  $IND = 4$ , la valeur de  $IND$  changeant à chaque changement de signe de  $DY - DE$ , ces boucles se faisant en modifiant  $XO$  par division du pas par 2 sans modifier le signe du pas si  $IND$  ne change pas de valeur, ou en changeant ce signe, si  $IND$  a changé de valeur.

Le travail en boucle se continue jusqu'à ce que l'erreur relative entre  $(DY - DE)$  et  $DE$  soit inférieure à  $.000001$  en valeur absolue, ou jusqu'à ce que le pas d'exploration devienne trop petit pour modifier effectivement la valeur de  $XO$  étant donné les possibilités de l'ordinateur en simple précision.

Dans les deux cas, nous admettons que les dernières valeurs calculées de  $XO$  et de  $DE$  le sont avec suffisamment de précision et nous calculons le paramètre  $S$  juste avant  $RETURN$ .

Si  $JND$  a pris la valeur 2 ou 3, la marche du calcul est semblable à celle décrite ci-dessus, la seule différence étant dans le mode de calcul de  $DY$ .

12.5.4. — SUBROUTINE PEXGEF calcul des paramètres inconnus d'une distribution exponentielle généralisée non tronquée, dont la valeur du paramètre de forme est connue, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance chaque fois que cela est possible, sinon en remplaçant l'équation provenant de la dérivée par rapport à  $x_0$  soit par l'équation contenant le moment centré d'ordre deux (cas de  $\delta = 1$ ), soit par l'équation contenant la moyenne arithmétique, soit par l'équation contenant la moyenne logarithmique (cf. paragraphe 8.3.).

Le programme principal doit fournir les valeurs  $X$  de la variable sous forme d'un vecteur, la taille  $N$  de l'échantillon, la valeur limite  $XS$  acceptée *a priori* pour  $XO$  (par exemple  $XS = 0$ ,  $XS = 1$ . E50 ou  $XS = + 1$ . E50, valeur qui doit être inférieure à la valeur minimale observée dans l'échantillon si le paramètre d'échelle est positif, ou supérieure à la valeur maximale observée si le paramètre d'échelle est négatif), le signe  $SS = + 1$  ou  $- 1$  choisi pour le paramètre d'échelle et la valeur, avec son signe, du paramètre de forme.

Le sous-programme calcule les valeurs du paramètre de position  $XO = x_0$  (borne inférieure de l'intervalle de définition de la variable si le paramètre d'échelle est positif,  $XO \geq XS$  ou borne supérieure si le paramètre d'échelle est négatif  $XO \leq XS$ ), du paramètre d'échelle  $S = s$  avec son signe choisi à l'avance, et de la moyenne  $XM$  de l'échantillon.

Nous calculons d'abord la moyenne de l'échantillon, la valeur moyenne des carrés des observations et nous cherchons la valeur minimale (si  $SS = +1$ ) ou maximale (si  $SS = -1$ ) observée.

Si le paramètre de forme est égal à 1, nous allons à l'instruction 6 qui calcule S et XO puis à RETURN.

Si le paramètre de forme est inférieur ou supérieur à 1, nous choisissons une valeur de XO, calculons une valeur DY du paramètre de forme d'après l'équation provenant de la dérivée du maximum de vraisemblance par rapport à  $x_0$  ou d'après l'équation qui la remplace, comparons cette valeur DY à la valeur DE fournie par le programme principal pour savoir dans quel sens faire varier XO que l'on suppose déterminé avec assez de précision lorsque l'erreur relative entre DY et DE est inférieure à .00001.

Le paramètre d'échelle est alors calculé juste avant RETURN.

12.5.5. — SUBROUTINE PEXGEP calcul des paramètres inconnus d'une distribution exponentielle généralisée non tronquée, dont la valeur du paramètre de position est connue, adaptée à un échantillon de taille connue, sans troncature, par la méthode du maximum de vraisemblance.

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, la taille N de l'échantillon, la valeur du paramètre de position (valeur qui doit être inférieure à la valeur minimale observée si le paramètre d'échelle est positif, ou supérieure à la valeur maximale observée si le paramètre d'échelle est négatif) le signe DS = +1 ou -1 choisi pour le paramètre de forme et le signe SS = +1 ou -1 choisi pour le paramètre d'échelle.

Le sous-programme calcule les valeurs du paramètre de forme DE et du paramètre d'échelle S (avec les signes prédéterminés) et la valeur moyenne XM de l'échantillon.

On choisit une valeur initiale DE = DS, calcule une valeur DZ d'après  $DZ = SM/SD - SL$  que l'on compare à DE, pour résoudre l'équation :

$$DE = \delta = \frac{\sum \text{Log} |x_i - x_0| \times |x_i - x_0|^{1/\delta}}{\sum |x_i - x_0|^{1/\delta}} - \frac{1}{n} \sum \text{Log} |x_i - x_0|,$$

ce qui correspond à une inversion de fonction facile à réaliser par la méthode de la tangente simple.

Le paramètre d'échelle est calculé juste avant RETURN.

12.5.6. — FUNCTION FEXGET fonction de distribution de la loi exponentielle généralisée tronquée.

12.5.7. — FUNCTION VEXGET inversion de la fonction exponentielle généralisée tronquée.

12.5.8. — SUBROUTINE PEXGET calcul des paramètres inconnus d'une distribution exponentielle généralisée (dans le cas d'une loi de Goodrich) tronquée adaptée à un échantillon avec troncature (cf. paragraphe 8.5.). On suppose connue la valeur du paramètre de position et les paramètres de forme (positif), d'échelle (positif) et de tronquage seront déterminés, les deux premiers par la méthode du maximum de vraisemblance.

Le programme principal doit fournir la totalité des valeurs X de la variable (y compris les valeurs inférieures au seuil de troncature) sous forme d'un vecteur, la taille N de l'échantillon, la valeur du paramètre de position XO, la valeur XH du seuil de troncature (qui pourrait être pris à  $(XO + \epsilon)$  valeur de la plus petite des observations différente de XO).

Le sous-programme calcule les valeurs du paramètre de forme  $DE = \delta$  positif, du paramètre d'échelle  $S = s$  positif, et du paramètre de tronquage  $FO = F_0$  compris entre 0 et 1.

Nous calculons d'abord la valeur moyenne de  $\text{Log} (X(J) - XO)$  pour les NH observations supérieures ou égales à XH.

Puis nous calculons la valeur du paramètre de forme comme dans la sous-routine PEXGE, la valeur du paramètre d'échelle et celle du paramètre de tronquage juste avant RETURN.

12.5.9. — SUBROUTINE PEXGEJ. Calcul des paramètres d'une distribution de Goodrich adaptée à un échantillon avec troncature : on suppose connue la valeur du paramètre de position, et les paramètres de forme, d'échelle et de troncature seront déterminés, les deux premiers par la méthode du maximum de vraisemblance.

Ce sous-programme a été établi pour l'étude de pluviométries journalières.

Le programme principal doit fournir : la valeur du paramètre de position XO et du seuil de troncature XH (qui pourrait être pris égal à la valeur de la plus petite des observations différente de XO).

Les observations sont transportées avec INTEGER \* 2 IX, ID en COMMON IX (70,366), ID (70) — IX (M, J) étant la hauteur pluviométrique correspondant au jour J de l'année dont les deux derniers chiffres du millésime forment M — ID (M), étant un indicateur d'utilisation de l'année M, valant 1 si les observations sont complètes et sans commentaires ; — 1 si les observations de l'année sont totalement manquantes ou si elles ne sont pas conservées pour le calcul, et 0 si les observations sont utilisées malgré les commentaires (observations incomplètes, non quotidiennes, douteuses...).

Le sous-programme calcule les valeurs du paramètre de forme DE =  $\delta$  positif, du paramètre d'échelle S = s positif et du paramètre de tronquage FO =  $F_0$ , compris entre zéro et un.

Nous commençons par calculer dans le sous-programme : le nombre relatif PH des observations supérieures ou égales à XH ainsi que la moyenne SL de leurs logarithmes naturels (le nombre relatif PH est le nombre des observations  $\geq$  XH divisé par le nombre total d'observations possibles, ici somme des 365 ou 366 jours de chaque année utilisée).

La méthode de travail est très simple puisque nous trouvons une équation, implicite et transcendante, contenant  $\delta$  seul parmi les paramètres à déterminer. Nous réalisons cette inversion de fonction par la méthode de la tangente seule en partant d'une valeur  $\delta = +1$  (instruction n° 12 à 20 non comprise).

Dès que DE =  $\delta$  est déterminé, avec une erreur relative de .00001 en valeur absolue, nous calculons le paramètre d'échelle S (instruction n° 20) et le paramètre de tronquage FO juste avant RETURN.

## 12.6. — Distribution bêta incomplète.

12.6.1. — FUNCTION FBÊTA. Fonction de distribution de la loi bêta incomplète non tronquée.

12.6.2. — FUNCTION VBÊTA. Inversion de la distribution bêta incomplète non tronquée.

12.6.3. — SUBROUTINE PBETAM. Calcul des paramètres d'une distribution bêta incomplète, non tronquée, adaptée à un échantillon de taille connue, sans troncature, par la méthode des moments.

Le programme principal doit fournir les valeurs X de la variable sous forme d'un vecteur, la taille N de l'échantillon.

Le sous-programme calcule les valeurs des paramètres de position XO et XI, du paramètre d'échelle S, des paramètres de forme P et Q et de la moyenne XM de l'échantillon.

## 12.7. — Distributions reliées à la loi normale.

12.7.1. — FUNCTION FSTUD. Fonction de répartition de la loi de Student-Fisher.

La probabilité, au dépassement, de la valeur (positive) t, c'est-à-dire la probabilité pour que la valeur absolue de x soit supérieure à t s'écrit :

$$Q = 1 - \frac{1}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \int_{-t}^t \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx,$$

$\nu$  étant le nombre de degrés de liberté.

En faisant le changement de variable  $\theta = \text{arctg} \frac{t}{\sqrt{v}}$ ,

si  $v$  est impair :

$$Q = 1 - \frac{2}{\pi} \left[ \theta + \sin \theta \cos \theta \left( 1 + \frac{2}{3} \cos^2 \theta + \dots + \frac{2.4 \dots (v-3)}{1.3 \dots (v-2)} (\cos \theta)^{v-3} \right) \right].$$

si  $v$  est pair :

$$Q = 1 - \sin \theta \left( 1 + \frac{1}{2} \cos^2 \theta + \dots + \frac{1.3 \dots (v-3)}{2.4 \dots (v-2)} (\cos \theta)^{v-2} \right).$$

Le programme principal doit fournir la valeur (positive) de  $t$  et le nombre de degrés de liberté.

12.7.2. — **FUNCTION VSTUD.** Inversion de la distribution de Student-Fisher.

On calcule la valeur positive  $t$  pour laquelle la probabilité au dépassement de la valeur absolue de  $x$  est égale à la probabilité donnée.

Le programme principal doit fournir cette probabilité et le nombre de degrés de liberté.

12.7.3. — **FUNCTION FCHID.** Fonction de répartition de la loi de  $\chi^2$  de K. Pearson.

La probabilité, au dépassement, de la valeur (positive)  $\chi^2$ , c'est-à-dire la probabilité pour que la valeur positive de  $t$  soit supérieure à  $\chi^2$ , s'écrit :

$$Q = 1 - \frac{1}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} \int_0^{\chi^2} t^{\frac{v}{2}-1} e^{-\frac{t}{2}} dt,$$

$v$  étant le nombre de degrés de liberté, en faisant le changement de variable  $u = \chi^2/2$  si  $v$  est pair :

$$Q = e^{-u} \left[ 1 + \sum_{r=1}^{\frac{v-2}{2}} \frac{u^r}{r!} \right]$$

et si  $v$  est impair :

$$Q = 1 - \text{erf}(\sqrt{u}) + \frac{2}{\sqrt{\pi}} \sqrt{u} e^{-u} \sum_{r=1}^{\frac{v-2}{2}} \frac{u^{r-1}}{1. \frac{3}{2} \dots \frac{2r-1}{2}}$$

Le programme principal doit fournir la valeur (positive) de  $\chi^2$  et le nombre de degrés de liberté.

12.7.4. — **FUNCTION VCHID** Inversion de la distribution de  $\chi^2$  de K. Pearson.

On calcule la valeur  $\chi^2$  pour laquelle la probabilité au dépassement est égale à la probabilité donnée.

Le programme principal doit fournir cette probabilité et le nombre de degrés de liberté.

12.8. — **Fonction gamma complète et ses dérivées.**

12.8.1. — **Fonction Gamma** : cette fonction existe dans la bibliothèque des ordinateurs IBM 360. Nous donnons ci-dessous une formule de calcul approchée pour  $z$  compris entre 1 et 2 (erreur sur  $\Gamma(z)$  inférieure à .0000001).

$$\Gamma(z) = 1. + (z - 1.) * (-.57710166 + (z - 1) * (.98585399 + (z - 1) * (-.87642182 + (z - 1) * (.83282120 + (z - 1) * (-.56847290 + (z - 1) * (.25482049 + (z - 1) * -.05149930))))))$$

Si  $z$  est compris entre 0 et 1, il faut utiliser la formule  $\Gamma(z) = \frac{1}{z} * \Gamma(z + 1)$ .

Si  $z$  est supérieur à 2, il faut utiliser la formule de récurrence,  $n$  étant un entier :

$$\Gamma(z - n) = (z - n - 1) * \Gamma(z - n - 1)$$

jusqu'à trouver un argument  $(z - n - 1)$  compris entre 1 et 2.

12.8.2. — Fonction logarithme de Gamma : cette fonction existe aussi dans la bibliothèque des ordinateurs IBM 360. Nous donnons ci-dessous une formule de calcul pour  $z$  supérieur à 18 (calcul à faire en double précision) :

$$\text{Log } \Gamma(z) = (z - .5) * \text{Log } z + .5 * \text{Log } 2\pi - z + 1./z * (1./12 + 1./z^2 * (-1./360 + 1/z^2 * (1./1260 - 1/z^2/1680)))$$

Si  $z$  est inférieur à 18, il faut utiliser la formule de récurrence,  $n$  étant un entier :

$$\text{Log } \Gamma(z + n - 1) = \text{Log } \Gamma(z + n) - \text{Log } (z + n - 1)$$

jusqu'à trouver un argument  $(z + n)$  supérieur à 18.

On peut prendre pour  $2\pi$  la valeur : 710./113. qui donne  $2\pi$  avec une erreur relative inférieure à  $1/10^7$ .

12.8.3. — FUNCTION VGAMMA. Inversion de la fonction Gamma complète.

La fonction Gamma complète admet un minimum de valeur  $\Gamma_1 = .8856031944$  pour une valeur de l'argument  $z_1 = 1.461632145$ . Pour une valeur inférieure à  $\Gamma_1$  l'inversion n'est pas possible et pour une valeur supérieure à  $\Gamma_1$  l'inversion conduit à deux solutions : l'une inférieure à  $z_1$  et l'autre supérieure.

Le programme principal doit fournir la valeur de  $\Gamma$  et une valeur  $-1$  ou  $+1$  d'un indicateur qui permet le choix entre les solutions inférieure ou supérieure à  $z_1$ .

12.8.4. — FUNCTION PSI. Dérivée logarithmique de la fonction Gamma.

Nous calculons  $\psi(z)$  par quadrature pour  $z$  compris entre 1 et 2. Si  $z$  est compris entre 0 et 1 il faut utiliser la formule :

$$\psi(z) = \psi(z + 1) - \frac{1}{z}$$

et si  $z$  est supérieur à 2, il faut utiliser la formule de récurrence,  $n$  étant un entier :

$$\psi(z - n) = \psi(z - n - 1) + \frac{1}{z - n - 1}$$

jusqu'à trouver un argument  $(z - n - 1)$  compris entre 1 et 2.

L'expression de l'intégrale donnant  $\psi(z)$  est :

$$\psi(z) = -C + \int_0^1 \frac{1 - t^{z-1}}{1-t} dt$$

$C$  constante d'Euler.

La dérivée de  $\psi(z)$  par rapport à  $t$ , pour  $z > 1$  tend vers  $+1$  si  $t$  tend vers  $0$  et vers  $(z-1)$  si  $t$  tend vers  $1$ .

La dérivée seconde de  $\psi(z)$  par rapport à  $t$ , pour  $z > 1$  tend vers  $+1$  si  $t$  tend vers  $0$  et vers  $\frac{(z-1)(z-2)}{2}$  si  $t$  tend vers  $1$  et la représentation de la dérivée première est donnée par la figure 27.

La méthode de quadrature employée est celle de Gauss, dont le principe est le suivant :

$$\int_{-1}^{+1} f(x) dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right)$$

le terme négligé étant  $\frac{1}{135} f^{IV}(\theta) \quad -1 < \theta < +1$ .

Appliquée à un intervalle  $a, b$  divisé en  $n$  sous-intervalles dont les centres sont à  $\frac{2r-1}{2n}(b-a)$ ,  $r$  variant de  $1$  à  $n$ , la quadrature s'écrit :

$$\frac{1}{2n} \sum_{r=1}^n \left[ f\left(\frac{b-a}{2n} \left(2r-1 - \frac{1}{\sqrt{3}}\right)\right) + f\left(\frac{b-a}{2n} \left(2r-1 + \frac{1}{\sqrt{3}}\right)\right) \right]$$

L'expérience nous a conduit à prendre 250 intervalles.

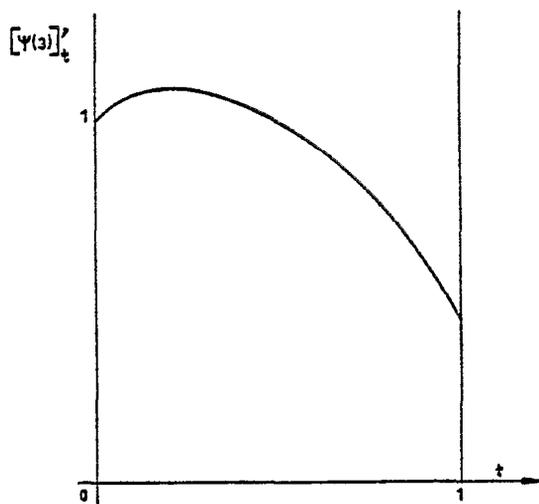


FIG. 27.

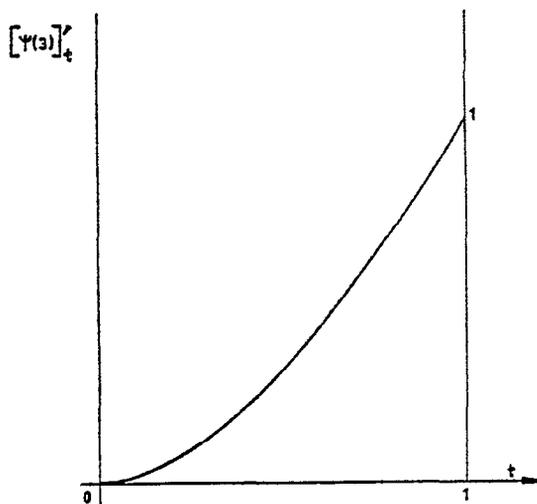


FIG. 28.

12.8.5. — **FONCTION VPSI.** Inversion de la fonction PSI qui est la dérivée logarithmique de la fonction Gamma complète.

La fonction  $\psi$  variant de  $-\infty$  et étant monotone croissante pour les valeurs positives de l'argument, l'inversion se fait facilement par la méthode de la tangente.

12.8.6. — **FONCTION PSIPR.** Dérivée de la fonction  $\psi$ .

Nous calculons  $\psi'(z)$  par quadrature pour  $z$  compris entre 2 et 3. Si  $z$  est inférieur à 2 et supérieur à 1 il faut utiliser la formule :

$$\psi'(z) = \psi'(z+1) + \frac{1}{z^2},$$

et si  $z$  est compris entre 0 et 1 :

$$\psi'(z) = \psi'(z+2) + \frac{1}{z^2} + \frac{1}{(z+1)^2},$$

Si  $z$  est supérieur à 3 il faut utiliser la formule de récurrence,  $n$  étant un entier :

$$\psi'(z-n) = \psi'(z-n-1) - \frac{1}{(z-n-1)^2},$$

jusqu'à trouver un argument  $(z-n-1)$  compris entre 2 et 3.

L'expression de l'intégrale donnant  $\psi'(z)$  est :

$$\psi'(z) = - \int_0^1 \frac{\text{Log } t \, t^{z-1}}{1-t} dt.$$

La dérivée de  $\psi'(z)$  par rapport à  $t$ , pour  $z > 2$ , tend vers 0 si  $t$  tend vers 0 et vers  $+1$  si  $t$  tend vers 1.

La dérivée seconde de  $\psi'(z)$  par rapport à  $t$ , pour  $z > 2$ , tend vers 0 si  $t$  tend vers 0 et vers  $(z-2)$  si  $t$  tend vers 1 et la représentation de la dérivée première est donnée par la figure 28.

La méthode de quadrature employée est celle de Gauss, avec 250 intervalles, comme pour la fonction  $\psi$ .

**ANNEXES**  
**LISTES FORTRAN**



```

SUBROUTINE PLOGNIX,S1,S,XO,XM,N,XS1
POH 009 SUBROUTINE PLOGNIX,FORME,ECHEL,POSIT,XMOY,N,VALLIM1
CALCUL DES PARAMETRES DE FORME,D ECHELLE ET DE POSITION,DE LA MOYENNE
D UNE LOI GAUSSO-LOGARITHMIQUE ADAPTEE A UN ECHANTILLON
DE N VALEURS FOURNIES EN VECTEUR PAR LE PROGRAMME PRINCIPAL
AINSI QUE LA VALEUR LIMITE INFERIEURE ACCEPTEE POUR POSIT
PAS DE SOUS-PROGRAMME A ADJOINDRE

DOUBLE PRECISION SL,SH,SLL,SLH,TL,TH
DIMENSION X(N)
XI=X(1)
XM=X(1)/N
DO 4 I=2,N
XM=XM+X(I)/N
4 XI=XM+X(I)/N
10 XI=XM+X(I)/N,XI
IND=1
AS=1
11 S=2,**(2.-IND)
12 AS=AS*S
XD=XI-1*XM-XI*AS
IF(XI-XD)100,100,17
13 IND=0
XD=XS
GO TO 17
14 AS=AS/2.
XA=XD
16 XD=XD+XM-XI*(3.-IND)*AS
IF(XA-XD)17,30,17
17 SLL=0
SH=0
SLL=0
SLH=0
DO 18 I=1,N
TH=1.00/(X(I)-XD)
TL=DLOG((X(I)-XD)*1.00)
SH=SH+TH/N
SL=SL+TL/N
SLL=SLL+TL*TL/N
18 SLH=SLH+TH*TL/N
SQ=SLL-SL*SL
SP=SL-SLH/SH
IF(IND)30,30,19
19 IF(FIX((SP/SQ-1.)*1.E5)122,30,20
20 IF(XO-XS)13,30,21
21 GO TO (12,24,24,14),IND
22 GO TO (23,14,12,25),IND
23 IND=3
GO TO 11
24 IND=4
GO TO 14
25 IND=2
GO TO 14
30 SI=SQRT(SQ)
S=DEXP(SL)
100 RETURN
END

```

```

SUBROUTINE PLOGNP(X,S1,S,XO,XM,N)
POH 010 SUBROUTINE PLOGNPX,FORME,ECHEL,POSIT,XMOY,N)
CALCUL DES PARAMETRES DE FORME ET D ECHELLE,DE LA MOYENNE
D UNE LOI GAUSSO-LOGARITHMIQUE ADAPTEE A UN ECHANTILLON
DE N VALEURS EN VECTEUR ET DE PARAMETRE DE POSITION (BORNE INF)
FOURNIS PAR LE PROGRAMME PRINCIPAL
PAS DE SOUS-PROGRAMME A ADJOINDRE

DOUBLE PRECISION SL,SLL,TL
DIMENSION X(N)
XM=0
SL=0
SLL=0
DO 1 I=1,N
XM=XM+X(I)/N
TL=DLOG((X(I)-XD)*1.00)
SL=SL+TL/N
1 SLL=SLL+TL*TL/N
SQ=SLL-SL*SL
SI=SQRT(SQ)
S=DEXP(SL)
RETURN
END

```

```

FUNCTION FLOGNT(X,S1,S,XO,FO)
POH 011 FUNCTION FLOGNT(X,FORME,ECHEL,POSIT,FO)
FUNCTION DE REPARTITION DE LA LOI GAUSSO-LOGARITHMIQUE TRONQUEE
CALCUL DE LA PROBABILITE CORRESPONDANT A LA VALEUR X DE LA VARIABLE
VALEUR X,PARAMETRES DE FORME,D ECHELLE ET DE POSITION (BORNE INF)
ET PARAMETRE DE TRONQUAGE FO
FOURNIS PAR LE PROGRAMME PRINCIPAL
PAS DE SOUS-PROGRAMME A ADJOINDRE

DOUBLE PRECISION U
U=DLOG((X-XO)/S*1.00)/SI*SQRT(.500)
FLOGNT=FO*(1.-FO)/2.*(1.+DERF(U))
RETURN
END

```

```

FUNCTION VLOGNT(P,S1,S,XO,FO)
POH 012 FUNCTION VLOGNT(FREDD,FORME,ECHEL,POSIT,FO)
INVERSION DE LA LOI GAUSSO-LOGARITHMIQUE TRONQUEE
CALCUL DE LA VARIABLE CORRESPONDANT A LA PROBABILITE FREDD
FREDD,PARAMETRES DE FORME,D ECHELLE ET DE POSITION (BORNE INF)
ET PARAMETRE DE TRONQUAGE FO
FOURNIS PAR LE PROGRAMME PRINCIPAL
PAS DE SOUS-PROGRAMME A ADJOINDRE

RED(U)=ALOG(U)/SI*.7071068
UI=0
IF(P-FO=.5E-7)20,20,4
4 EX=2.*P-FO/(1.-FO)-1.
C=1.
UI=EXP(-SI*SI)
EI=ERF(RED(U))
5 AS=1./2.*MC
IF(FIX((EX-EI)*1.E7)17,20,4
6 T=1.
GO TO 8
7 T=-1.
8 UI=(F(1.+T*AS)
EI=ERF(RED(U))
IF(FIX((EX-EI)*1.E7)*T)9,20,8
9 C=C*1.
IF(C-19.15,10,11
10 UO=UI
EO=EX-EI
GO TO 5
11 EI=EI*SI
UI=UO-EO*(UO-UI)/(EO-EI)
20 VLOGNT=S*UI+XO
RETURN
END.

```

```

SUBROUTINE PLOGNT(X,S1,S,XO,FO,N,SEUIL)
POH 013 SUBROUTINE PLOGNTX,FORME,ECHEL,POSIT,FO,N,SEUIL)
CALCUL DES PARAMETRES DE FORME,D ECHELLE ET DE TRONQUAGE
D UNE LOI GAUSSO-LOGARITHMIQUE ADAPTEE A UN ECHANTILLON
DE N VALEURS EN VECTEUR ET DE PARAMETRE DE POSITION (BORNE INF)
FOURNIS PAR LE PROGRAMME PRINCIPAL
AINSI QUE LA VALEUR DU SEUIL DE TRONCATURE
ADJOINDRE SOUS-PROGRAMME FUNCTION FLOGNT

DIMENSION X(N)
DOUBLE PRECISION Y,TL,TH
FR=0
TL=0
TH=0
NM=0
PI=2.506628
UL=ALOG(XM-XO)
DO 5 J=1,N
IF(X(J)-XM)5,5,4
4 T=DLOG((X(J)-XD)*1.00)
TL=TL+T
TH=TH+T*T
NM=NM+1
5 CONTINUE
SL=TL/NM
SM=TH/NM
FM=NM/(NM*1.)
SI=2.
JND=1
10 AS=1./2.
GO TO 13
11 JND=3
AS=2.
12 SA=SI
13 SI=SI*AS
14 SS=(SM-SL*UL-SI*SI)/(SL-UL)
S=EXP(SS)
GC=(UL-SS)/SI
XL=SS*SI*EXP(-GC*GC/2.)/PI/TL.-FLOGNT(XM,S1,S,XO,FR)
IF(FIX((XL/SL-1.)*1.E5)15,25,16
15 GO TO (15,18,17),JND
16 GO TO (11,20,12),JND
17 JND=2
18 SB=SI
19 SI=(SA+SB)/2.
IF((SI-SA)*(SI-SB))14,25,25
20 SA=SI
GO TO 19
25 FM=MAX(10.,(1.-FM)/(1.-FLOGNT(XM,S1,S,XO,FR)))
RETURN
END

```





C  
C  
C  
C  
C  
C  
C  
C  
C  
C

FUNCTION FEXGE(X,DE,S,XO)

PGH 022 FUNCTION FEXGE(X,FORME,ECHEL,POSIT)

FNCTION DE REPARTITION DE LA LOI EXPONENTIELLE GENERALISEE  
CALCUL DE LA PROBABILITE CORRESPONDANT A LA VALEUR X DE LA VARIABLE  
VALEUR X,PARAMETRES DE FORME,D ECHELLE ET DE POSITION (BORNE)  
FOURNIS PAR LE PROGRAMME PRINCIPAL  
PAS DE SOUS-PROGRAMME A ADJOINDRE

DOUBLE PRECISION U  
SD=SIGN(1.,(S\*DE))  
U=(X-XO)/S  
U=U\*(1.00/DE)  
FEXGE=.5\*(SD+1.-SD\*EXP(-U))  
RETURN  
END

FUNCTION VEXGE(F,DE,S,XO)

PGH 023 FUNCTION VEXGE(FREDD,FORME,ECHEL,POSIT)

INVERSION DE LA LOI EXPONENTIELLE GENERALISEE  
CALCUL DE LA VARIABLE CORRESPONDANT A LA PROBABILITE FREDD  
FREDD,PARAMETRES DE FORME,D ECHELLE ET DE POSITION (BORNE)  
FOURNIS PAR LE PROGRAMME PRINCIPAL  
PAS DE SOUS-PROGRAMME A ADJOINDRE

DOUBLE PRECISION F  
SD=SIGN(1.,(S\*DE))  
F=.500\*(SD+1.-SD\*\*F1.00)  
F=DLOG(F)  
U=(F)\*\*4\*DE\*(1.00)  
VEXGE=S\*U\*XO  
RETURN  
END

C  
C  
C  
C  
C  
C  
C  
C  
C  
C

SUBROUTINE PEXGE(X,DE,S,XO,XM,N,XS,SS,DS)

PGH 024 SUBROUTINE PEXGE(X,FORME,ECHEL,POSIT,XMOY,N,VALLIM,STIECH,STFOR)

CALCUL DES PARAMETRES DE FORME,D ECHELLE ET DE POSITION,DE LA MOYENNE  
D UNE LOI EXPONENTIELLE GENERALISEE ADAPTEE A UN ECHANTILLON  
DE N VALEURS FOURNIES EN VECTEUR PAR LE PROGRAMME PRINCIPAL  
AINSI QUE LA VALEUR LIMITE ACCEPTEE POUR POSIT ET LES SIGNES DES  
PARAMETRES D ECHELLE ET DE FORME  
ADJOINDRE SOUS-PROGRAMME FUNCTION VGHANNA

DOUBLE PRECISION A,B,C,SD,SE,SL,SH,SM,SN  
DIMENSION X(N)  
XI=X(1)/SS  
XM=X(1)/N  
DS=1./N  
DO 5 I=2,N  
XM=XM+X(I)/N  
5 XI=AMINI(X(I)/SS,XI)  
XI=XI\*SS  
JND=0  
7 AS=1.  
DE=OS/2.  
JND=JND+1  
IND=1  
10 AS=AS/2.  
GO TO 13  
11 IND=3  
12 AS=AS\*2.  
13 XO=XI-(XM-XI)\*AS  
IF(IXI-XO)/SS)7,18  
14 IND=0  
XO=XI  
GO TO 18  
15 AS=AS/2.  
XA=XO  
17 XO=XO+(XM-XI)\*(3.-(IND)\*AS  
IF(XA-XO)/18,50,18  
18 KND=1  
19 SD=0  
SE=0  
SL=0  
SH=0  
SM=0  
SN=0  
DO 20 I=1,N  
A=SS\*(X(I)-AO)  
B=DLOG(A)  
C=A\*(1./DO/DE)  
SD=SD+C/N  
SE=SE+C/A/N  
SL=SL+B/N  
SH=SH+1./A/N  
SM=SM+B\*C/N  
SN=SM+B\*C/N  
20 DZ=SM/SD-SL  
IF(IXI(DZ/DE-1.)\*1.E5)121,32,22  
21 GO TO (23,26,27),KND  
22 GO TO (24,25,29),KND  
23 DE=DE/2.  
GO TO 19  
24 KND=2  
25 DA=DE  
DE=DE\*2.  
GO TO 19  
26 KND=3  
27 DB=DE  
28 DE=(DA+DB)/2.  
GO TO 19  
29 DA=DE  
GO TO 28  
32 IF(IND)50,50,34  
34 LND=AMINI(3.,AMAX(1.,(DE+1)))  
LND=MAXO(LND,JND)  
GO TO (35,36,37),LND  
35 DY=1.-SE/SH/SD  
GO TO 39  
36 DY=SS\*(XM-XO)/SD\*\*DE  
DY=VGHANNA(DY,(LND-1))-1.  
GO TO 39  
37 DY=SL/(DLOG(SD)-.5772156649)  
39 IF(IXI(DY/DE-1.)\*1.E6)142,50,40  
40 IF(SS\*(XO-XS))14,50,41  
41 GO TO (11,44,12,13),IND  
42 GO TO (10,15,43,43),IND  
43 IND=2  
GO TO 15  
44 IND=4  
GO TO 15  
50 S=SS\*SD\*\*DE  
100 RETURN  
END







