

Recherche d'un test d'ajustement

Y. BRUNET-MORET

Ingénieur Hydrologue au Bureau Central Hydrologique
de l'ORSTOM à Paris

RÉSUMÉ

Présentation d'un test d'ajustement, valable pour des échantillons de petites tailles, établi d'une façon expérimentale en utilisant les observations rangées et en faisant référence à la méthode du maximum de vraisemblance pour la détermination des valeurs numériques des paramètres.

ABSTRACT

Exposing a test of fit (composite hypothesis) available with short samples, experimentally established, using the order — statistics (observations ordered) and referring to maximum likelihood estimators for parameters.

SOMMAIRE

1. *Introduction, conditions proposées*
2. *Estimation des paramètres*
3. *Rappel au sujet des tests existants*
4. *Principe du test proposé*
5. *L'échantillon « idéal »*
6. *Essais du test*
7. *Distribution du test*
8. *Utilisation du test, conclusion*
9. *Annexes*

1. INTRODUCTION : CONDITIONS PROPOSÉES

1.1. Le motif de cette recherche est de proposer un test d'ajustement à but essentiellement pratique qui, pour un échantillon de petite taille de variables aléatoires absolument continues, permette de choisir parmi les multiples fonctions de répartition utilisables, celle qui semble le moins mal représenter l'échantillon : les fonctions de répartition étant de formulations mathématiques choisies a priori, mais les valeurs numériques des paramètres étant calculées d'après l'échantillon.

1.1.1. Une qualité exigée du test est de retrouver, en cherchant sa valeur minimale par variation des valeurs numériques des paramètres, les mêmes valeurs de ces paramètres que celle que donnerait leur estimation par une autre méthode que l'on sait par ailleurs être la meilleure. La meilleure méthode de calcul des paramètres doit alors conduire à la valeur minimale du test.

1.1.2. Une autre qualité est de pouvoir ne tester la valeur de l'ajustement que sur une partie de l'échantillon, vers les valeurs les plus fortes ou les plus faibles, pour mieux s'assurer de la consistance de l'extrapolation vers les valeurs les plus fortes ou les plus faibles.

1.1.3. Une condition nécessaire à remplir pour que l'utilisation du test soit facile, est qu'il soit non paramétrique : c'est-à-dire que sa signification soit indépendante tant de la formulation mathématique de la fonction de répartition que des valeurs numériques des paramètres.

2. ESTIMATION DES PARAMÈTRES

La première question qui se pose est celle de « la meilleure » estimation des paramètres.

2.1. DÉFINITION CLASSIQUE DE L'ESTIMATEUR SANS BIAIS

2.1.1. Soient un échantillon de n variables aléatoires et indépendantes $x_1 \dots x_n$ et une fonction de répartition choisie a priori (continue) $F(x, \theta_1, \dots, \theta_k)$ dont les paramètres $\theta_1 \dots \theta_k$ n'ont pas des valeurs connues ou choisies a priori mais à estimer d'après l'échantillon.

L'estimateur de θ_i est : $t_i = g_i(x_1 \dots x_n, n)$.

La forme mathématique de la fonction g_i dépend d'une part de celle de la fonction de répartition F , d'autre part de la méthode d'estimation : maximum de vraisemblance, moments...

La valeur numérique de l'estimation dépend des valeurs numériques des variables de l'échantillon et de la taille de ce dernier.

2.1.2. En supposant l'estimateur convergent (correct) et en considérant qu'on pourrait tirer, d'une façon aléatoire, tous les échantillons possibles de taille n de la population mère, l'estimateur t_i doit être considéré comme une variable aléatoire qui suit une fonction de répartition H_i et possédant une moyenne $E(t_i)$.

D'après la définition classique l'estimateur est sans biais lorsque, quel que soit n

$$E(t_i) = \theta_i \text{ véritable valeur du paramètre dans la population mère}$$

La définition s'étend à tout estimateur, par exemple à celui de la valeur de la variable associée à une probabilité choisie à l'avance.

2.2. CRITIQUE DE LA DÉFINITION

2.2.1. Le mieux est de citer KENDALL (t. II, chap. 17, p. 4).

« Considérons la distribution d'échantillonnage d'un estimateur t . Si l'estimateur est correct, cette distribution aura, pour des échantillons de grande taille n , une valeur centrale dans le voisinage de la valeur θ du paramètre de la population mère. On peut choisir parmi les estimateurs consistants celui qui répond à la condition que θ soit égal à la valeur centrale ci-dessus non seulement pour n grand, mais quel que soit n . Et on définit ce qu'on appelle l'estimateur non biaisé par la relation $E(t) = \theta$.

L'expression « non biaisé » est malheureusement employée comme beaucoup d'autres en statistique. Il n'y a pas d'autre motif que celui de commodité, pour privilégier la moyenne arithmétique parmi les autres mesures de valeurs centrales comme critère de biais. On aurait pu aussi bien choisir la médiane de la distribution de t , ou son mode, comme déterminant l'estimateur non biaisé. La moyenne est utilisée, comme toujours, pour sa commodité mathématique. C'est légitime, mais il est nécessaire de spécifier que le terme « non biaisé » ne doit pas être pris comme impliquant des propriétés d'une nature autre qu'arithmétique ».

2.2.2. La définition de l'estimateur sans biais fait référence à la valeur, inconnue, du paramètre dans la population mère. De plus $E(t)$ devrait être calculé connaissant tous les échantillons possibles de taille n pouvant être tirés de la population mère. Or, nous n'avons à notre disposition qu'un seul échantillon.

2.3. CHOIX DE L'ESTIMATEUR DU « MAXIMUM DE VRAISEMBLANCE »

2.3.1. Soient un échantillon de n variables aléatoires et indépendantes et une fonction de répartition choisie a priori (continue) $F(x, \theta_1 \dots \theta_k)$ dont les paramètres $\theta_1 \dots \theta_k$ n'ont pas de valeurs connues ou choisies a priori, mais à estimer d'après l'échantillon.

En considérant les paramètres comme variables aléatoires (continues, dans leurs domaines de définition) on peut associer à des valeurs $\tau_1 \dots \tau_k$ choisies pour ces paramètres une densité de probabilité qui est celle d'observer l'échantillon donné $x_1 \dots x_n$ par tirage au hasard dans la population mère définie par $F(x, \tau_1 \dots \tau_k)$: on définit la vraisemblance de l'échantillon comme proportionnelle au produit des densités de probabilités $dF(x_i)/dx$.

2.3.2. Il semble logique, étant donné qu'on fait un « pari » sur un seul tirage, de choisir l'estimateur qui conduit à l'estimation la plus probable d'après l'échantillon (et la fonction de répartition choisie).

L'emploi de la méthode du maximum de vraisemblance calcule les valeurs $\tau_1 \dots \tau_k$ qui maximisent leur densité de probabilité : valeurs modales de leur distribution conjointe (et non valeurs modales des distributions marginales des paramètres).

C'est pourquoi nous considérons que les valeurs des paramètres calculées par le maximum de vraisemblance sont « les meilleures » (1.2).

2.4. AVANTAGES DE LA MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

Outre les avantages de cette méthode, nous pouvons ajouter :

2.4.1. La valeur x_p correspondant à une probabilité, fixée a priori, P dans la fonction de distribution $F(x, x_0, s, \alpha, \beta \dots)$ est la valeur la plus vraisemblable (modale) si les paramètres x_0 de position, s d'échelle, $\alpha, \beta \dots$ de forme sont calculés par la méthode du maximum de vraisemblance

$$P = \int_{-\infty}^{x_p} f(x, x_0, s, \alpha, \beta \dots) dx = \int_{-\infty}^{u_p} f(u, \alpha, \beta \dots) du$$

$$\text{avec } u = \frac{x - x_0}{s} \quad x_p = su_p + x_0$$

u_p n'étant fonction ni de x_0 ni de s , mais étant fonction des paramètres de forme (et de la valeur de la probabilité P fixée) : $u_p = \Phi(\alpha, \beta \dots)$.

En prenant x_p comme paramètre de position, les équations du maximum de vraisemblance s'écrivent :

$$\frac{\partial \mathcal{L}}{\partial x_p} = 0 = \frac{\partial \mathcal{L}}{\partial x_0} + \Phi(\alpha, \beta \dots) \frac{\partial \mathcal{L}}{\partial s} + s \Phi'_{\alpha} \frac{\partial \mathcal{L}}{\partial \alpha} + s \Phi'_{\beta} \frac{\partial \mathcal{L}}{\partial \beta} + \dots$$

$$\frac{\partial \mathcal{L}}{\partial s} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \text{ etc.}$$

La première équation se réduit à $\frac{\partial \mathcal{L}}{\partial x_0} = 0$ et on retrouve le système d'équations de calcul des paramètres $x_0, s, \alpha, \beta \dots$ par le maximum de vraisemblance.

Par contre l'utilisation d'estimateurs classiquement « sans biais » ne conduit pas à une estimation « sans biais » de x_p : sauf cas très particulier

$$E(u_p) \text{ n'est pas égal à } \Phi(E[\alpha], E[\beta] \dots)$$

$$E(x_p) \text{ n'est pas égal à } E(x_0) + E(s) \cdot E(u_p),$$

car les distributions des paramètres ne sont pas indépendantes (sauf dans le cas de la loi normale).

2.4.2. La valeur P de la probabilité correspondant à une valeur x_p fixée a priori est la valeur la plus vraisemblable (modale) si les paramètres de la distribution sont calculés par le maximum de vraisemblance.

En partant de $P = F(x_p, x_0, s, \alpha, \beta \dots)$

$$\frac{\partial \mathcal{L}}{\partial P} = F'_{x_0} \frac{\partial \mathcal{L}}{\partial x_0} + F'_s \frac{\partial \mathcal{L}}{\partial s} + F'_\alpha \frac{\partial \mathcal{L}}{\partial \alpha} + \dots$$

d'où $\frac{\partial \mathcal{L}}{\partial P} = 0$.

Par contre, sauf cas très particulier :

$$E(P) \text{ n'est pas égal à } F(x_p, E(x_0), E(s), E(\alpha) \dots)$$

3. RAPPEL AU SUJET DES TESTS EXISTANTS

3.1. CAS D'UTILISATION

Il est nécessaire d'insister tout d'abord sur les deux façons d'utiliser les tests d'ajustement :

3.1.1. Cas A :

Hypothèse simple. La loi de distribution de la population mère est parfaitement connue a priori (par sa formulation mathématique et les valeurs numériques des paramètres). On teste la probabilité de l'échantillon comme représentation de la population mère, ou, dit autrement, la probabilité du tirage au hasard de cet échantillon dans la population mère parfaitement connue.

On est aussi dans ce cas lorsque la distribution de la population mère n'est pas connue, mais entièrement choisie à l'avance : formulation mathématique et valeurs numériques des paramètres.

3.1.2. Cas B :

Hypothèse complexe. La formulation mathématique de la loi de distribution de la population mère est choisie à l'avance (ou connue a priori), mais les valeurs numériques de tous ou certains paramètres (au moins un) sont calculées d'après l'échantillon disponible. On teste la valeur de la loi de distribution choisie (avec ses paramètres calculés) comme représentation de l'échantillon, ou, dit autrement, la probabilité du tirage au hasard de cet échantillon dans la distribution de la population mère qui en est déduite compte tenu du choix de la formulation mathématique et du mode de calcul des valeurs numériques des paramètres.

3.2. TESTS EXISTANTS

3.2.1. Pour le cas A, dans lequel à la vérité on ne cherche pas à vérifier la validité d'un ajustement, il existe

un certain nombre de tests de bonne puissance : Cramer, Anderson, Kolmogorov... qui sont « non paramétriques », c'est-à-dire que leur puissance et leur signification sont indépendantes de la formulation mathématique de la loi de distribution de la population mère comme des valeurs numériques, connues, des paramètres entrant dans cette distribution.

Pour mémoire, on peut signaler les tests paramétriques bien connus, applicables lorsque la population mère suit une loi normale, de Student et de Fisher.

Le test du χ^2 a été tout d'abord conçu pour ce cas A. Il n'est valable qu'avec un échantillon de taille assez grande et sa puissance est bien inférieure aux puissances des tests cités en premier lieu : par exemple 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 seuil de signification du test du χ^2 sera de l'ordre de $N^{5/4}$ ($N = 120$, $N^{5/4} = 400$).

3.2.2. Pour le cas B, il n'existe que le test du χ^2 qui ne vaut pas grand chose. Entre autres défauts : même dans le cas d'échantillons assez grands, la valeur du test du χ^2 renseigne seulement sur la possibilité qu'a la loi choisie (avec ses paramètres calculés) de représenter la distribution de l'échantillon observé dans sa zone de forte densité de probabilité.

Le test que nous avons recherché correspond à ce cas B, est utilisable pour de petits échantillons, et semble être quasi non-paramétrique.

3.3. NIVEAU DE SIGNIFICATION ET PUISSANCE D'UN TEST

3.3.1. L'hypothèse testée peut être vraie ou fausse, qu'elle soit simple ou qu'elle soit complexe. Il y a deux erreurs possibles :

A : on peut rejeter à tort l'hypothèse si elle est vraie

B : on peut l'accepter à tort si elle est fausse.

La probabilité de A est le niveau de signification « α », probabilité au dépassement définie par la valeur numérique du test.

La probabilité de B est « β » et la puissance du test est $1-\beta$, qui dépend de la formulation du test, de la taille de l'échantillon et du niveau de signification α .

3.3.2. L'habitude semble prise de rejeter l'hypothèse si le niveau de signification est $\leq 1\%$, de l'accepter si ce niveau est $\geq 5\%$.

Il faut bien voir qu'en moyenne des échantillons tirés d'une population mère de formulation mathématique connue a priori (que les valeurs numériques des paramètres soient connues a priori ou qu'elles soient calculées d'après chaque échantillon, indépendamment les uns des autres) fourniront, pour le niveau de signification, une valeur moyenne de 0,5. En effet, dans les conditions ci-dessus, la distribution de ce niveau est continue et uniforme entre 0 et 1.

Cette remarque doit être utilisée chaque fois qu'on pense que toute une série de phénomènes suit la même loi de distribution avec des valeurs numériques des paramètres différentes pour chaque terme : on suppose que chaque terme de la série est représenté par un échantillon, les échantillons étant indépendants entre eux (sans corrélation) et constitués de variables aléatoires et indépendantes entre elles. Ces conditions théoriques sont très restrictives : mais dans la pratique on peut admettre des échantillons corrélés entre eux et autocorrélés à condition que les coefficients de corrélation et d'autocorrélation ne soient pas trop élevés (en valeurs absolues).

4. PRINCIPE DU TEST PROPOSÉ

4.1. Une des idées directrices pour l'établissement du test étant qu'il soit non paramétrique, il semble nécessaire de passer (comme pour le test de χ^2) par l'intermédiaire des fréquences calculées d'après les valeurs de l'échantillon, la formulation mathématique de la loi de distribution choisie, et les valeurs numériques estimées pour les paramètres.

A chaque variable de l'échantillon de taille n rangé en ordre croissant correspond une fréquence (au non dépassement) F_i et par inversion de la loi normale centrée réduite, une valeur v_i

$$F_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v_i} e^{-\frac{v^2}{2}} dv$$

4.2. On peut concevoir un échantillon idéal de taille n , de valeurs u_i rangées en ordre croissant et suivant « exactement » une loi normale, en obéissant aux conditions $\sum u_i = 0$, $\sum u_i^2 = n$ (détermination de la moyenne et de la variance par la méthode du maximum de vraisemblance).

Les quantités $(v_i - u_i)$ sont une mesure de la divergence entre l'échantillon des v et l'échantillon idéal des u . Plutôt que de prendre $\sum_1^n (v_i - u_i)$ comme valeur du test, il a semblé préférable de prendre une somme de carrés de surfaces comprises entre un axe d'abscisses équidistantes et la ligne brisée définie par les points Z_i d'abscisses $2i$ et d'ordonnées $(v_i - u_i)$.

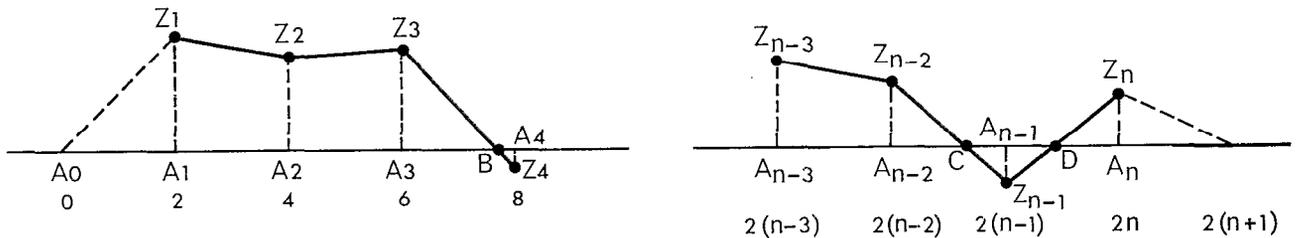


Figure 1

La figure 1 ci-dessus explicite les surfaces élémentaires prises en compte et élevées au carré :

pour le premier point :	A_0	Z_1	A_2
deuxième	A_1	Z_2	A_3
troisième	A_2	Z_3	B
$(n - 2)^{\text{ième}}$	A_{n-3}	Z_{n-2}	C
$(n - 1)^{\text{ième}}$	C	Z_{n-1}	D
$n^{\text{ième}}$	D	Z_n	A_{n+1}

On voit que lorsque la ligne brisée joignant les Z_i coupe l'axe des abscisses, les surfaces élémentaires prises en compte sont bien diminuées par rapport à celles prises en compte lorsque la ligne brisée ne coupe pas l'axe des abscisses.

4.3. Des essais de simulation montrent que les variances des quantités $(v_i - u_i)$ sont minimales au rang $n/2$ et croissent fortement vers chaque extrémité (par exemple pour $n = 39$ les variances correspondant aux rangs 1 et n sont dix fois plus fortes que les variances correspondant aux rangs 14 à 26). A mêmes probabilités d'apparition de quantités $(V-U)$ les carrés des surfaces élémentaires correspondantes sont dans le rapport des variances. Il s'ensuit que, dans le calcul de la valeur du test, le poids des valeurs extrêmes est bien supérieur au poids des valeurs centrales.

5. L'ÉCHANTILLON « IDÉAL »

L'échantillon « idéal » de valeurs U de référence ne semble pas facile à définir d'une façon mathématique stricte. C'est le point délicat de la construction du test, il n'a pu être résolu que d'une façon expérimentale.

5.1. Il a été effectué des tirages au hasard d'échantillons de taille n dans une distribution normale. Chaque échantillon a été rangé et recalculé en variables X , centrées réduites.

Le domaine de variation de chaque variable x_i est défini entre les limites suivantes :

$$\begin{array}{lll} \text{rang } i = 1 & \text{limite inférieure } -\sqrt{n-1} & \text{limite supérieure } -\frac{1}{\sqrt{n-1}} \\ 1 < i < n & -\sqrt{\frac{n-i}{i}} & +\sqrt{\frac{i-1}{n-i+1}} \\ i = n & +\frac{1}{\sqrt{n-1}} & +\sqrt{n-1} \end{array}$$

et les variables x_i sont liées entre elles par les conditions

$$\sum x_i = 0 \quad \sum x_i^2 = n$$

$x_1 \leq x_2 \dots \leq x_{n-1} \leq x_n$ dont au moins une inégalité.

5.1.1. Pour l'ensemble des échantillons de taille n , il a été calculé pour chaque rang : la valeur moyenne \bar{x}_i (parce que c'est la seule valeur centrale facile à calculer), la variance w_i et le coefficient d'asymétrie γ_i .

On vérifie expérimentalement (compte tenu des écarts dus aux tirages au hasard)

$$\bar{x}_i = -\bar{x}_{n+1-i} \quad w_i = w_{n+1-i} \quad \gamma_i = -\gamma_{n+1-i}$$

$$n = \sum_1^n \bar{x}_i^2 + \sum_1^n w_i$$

$$\sum_1^n |x_i| \text{ semble bien être } \frac{2n}{\sqrt{2\pi}} \text{ soit } n \int_{-\infty}^{+\infty} \frac{|t|}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

La somme $\sum_1^n w_i$ peut s'écrire avec une bonne approximation de $n = 2$ à $n = 200$ (fig. 2)

$$\sum w_i = \left(0,204 - \frac{1}{1,8(n+1,8)^{1,5}} \right) \text{Log} \left(\frac{n-1,8}{0,2} \right)$$

5.1.2. D'autre part la variance est minimale pour $i = \frac{n+1}{2}$ (ou les valeurs entières les plus voisines si n est pair) et de valeur approximative $w_0 = 0,5642 \frac{n}{n^2+1}$ et elle est maximale pour $i = 1$ ou n , de valeur approximative :

$$w_1 = w_n = \frac{n-2}{4} w_0.$$

Cette variance pour le rang 2 est presque indépendante de n , d'une valeur voisine de 0,06.

Du rang immédiatement supérieur à $\frac{n+1}{4}$ au rang immédiatement inférieur à $(n+1) - (n+1)/4$, c'est-à-dire pour la moitié, les valeurs centrales, de l'échantillon rangé, la variance est très voisine de w_0 (fig. 3, 4 5 et 6).

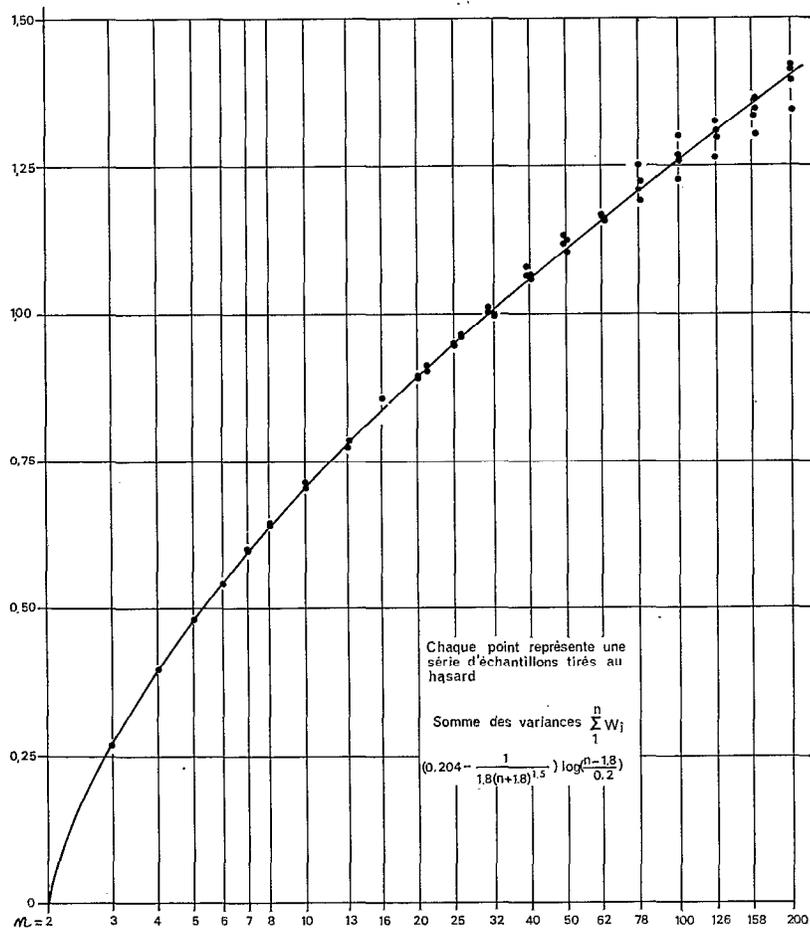


Figure 2

5.2. Nous admettons que l'échantillon des valeurs \bar{x}_i est en rapport étroit, inconnu, avec l'échantillon idéal de valeurs u .

5.2.1. Les \bar{x}_i n'étant pas faciles à calculer, il a été choisi de les remplacer par les valeurs y_i définies par

$$\frac{i - 0,3}{n + 0,4} = \int_{-\infty}^{y_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Les y_i sont liés aux \bar{x}_i par un facteur multiplicateur quasi constant quel que soit i pour une taille n donnée : $a_i y_i = \bar{x}_i$, $a_i > 1$ (fig. 3, 4, 5 et 6).

5.2.2. Nous admettons que l'échantillon idéal des valeurs u s'obtient, avec une approximation suffisante, en multipliant les valeurs y_i par une quantité c :

$$c = \sqrt{\frac{n + \sum w_i}{\sum y_i^2}}, \quad u_i = c y_i$$

La somme des carrés des valeurs u est alors égale à $n +$ somme des variances des \bar{x}_i .

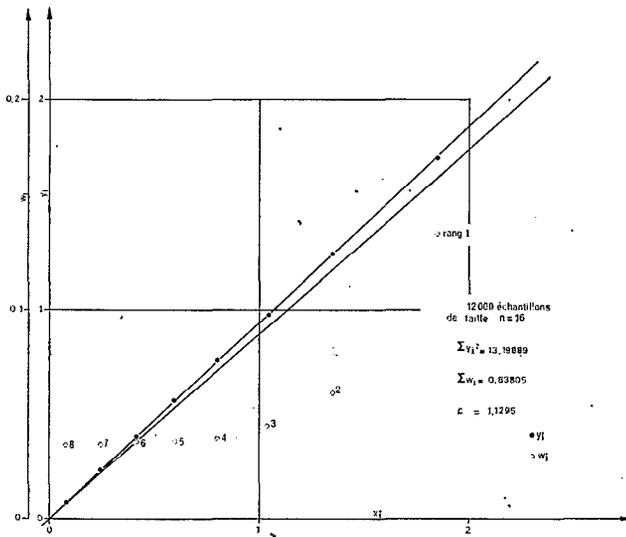


Figure 3

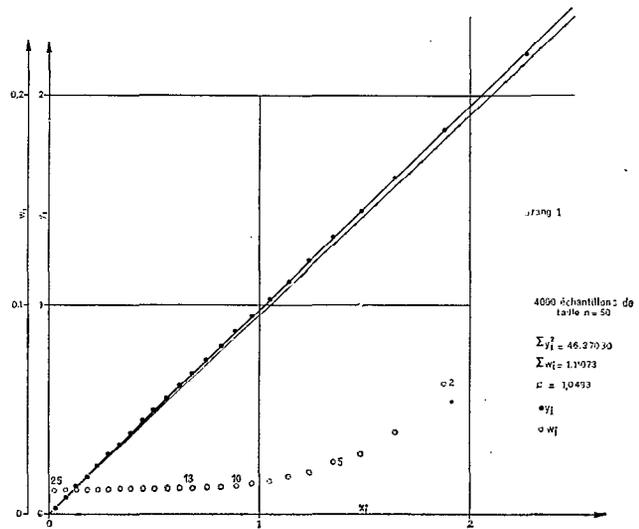


Figure 4

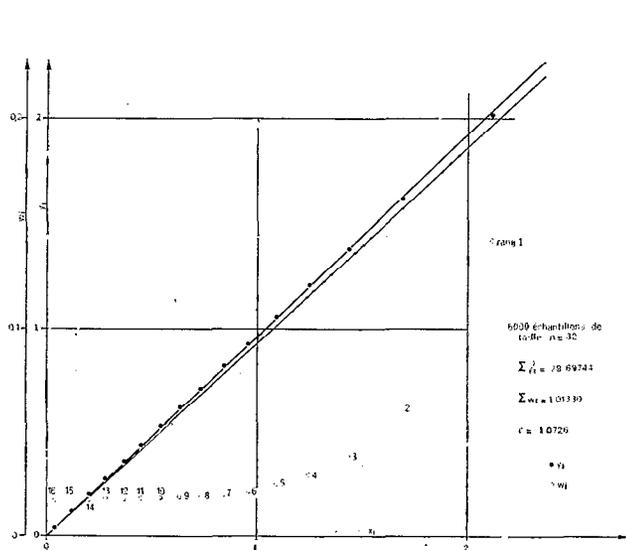


Figure 5

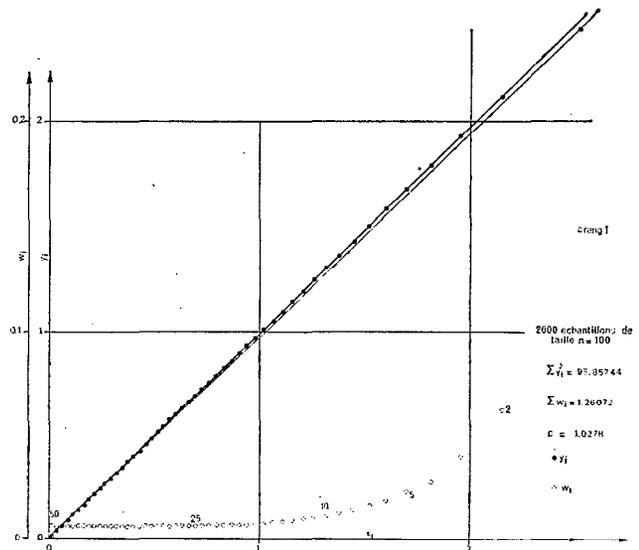


Figure 6

6. ESSAIS DU TEST

Il a été vérifié par quelques tirages au hasard que les qualités exigées en 1.1.1 et 1.1.3 étaient convenablement obtenues du test (qui par construction possède la qualité 1.1.2).

6.1. ESSAIS EN DISTRIBUTION NORMALE

Population mère distribuée normalement, paramètres de position nul, d'échelle égal à 1.

Nous appelons x_0 et s_0 les valeurs des paramètres de position et d'échelle déterminés sur un échantillon de taille n par la méthode du maximum de vraisemblance, x_m et s_m les valeurs des paramètres déterminés en minimisant la valeur du test, \bar{x}_0 et \bar{s}_0 , \bar{x}_m et \bar{s}_m les valeurs moyennes de ces paramètres sur ℓ échantillons de taille n .

Les écarts-types (approchés par défaut) sont pour \bar{x}_0 : $\frac{1}{\sqrt{n \ell}}$, pour \bar{s}_0 : $\frac{1}{\sqrt{2 n \ell}}$.

6.1.1. $\ell = 100$ échantillons de taille $n = 5$

\bar{x}_m est à $-0,0415$ écart-type de \bar{x}_0

s_m est à $-1,022$ écart-type de s_0

6.1.2. $\ell = 100$ échantillons de taille $n = 10$

\bar{x}_m est à $+0,0058$ écart-type de \bar{x}_0

s_m est à $-0,6644$ écart-type de s_0

6.1.3. $\ell = 100$ échantillons de taille $n = 25$

\bar{x}_m est à $+0,0463$ écart-type de \bar{x}_0

s_m est à $-0,1616$ écart-type de s_0

6.1.4. $\ell = 100$ échantillons de taille $n = 50$

\bar{x}_m est à $+0,0258$ écart-type de \bar{x}_0

\bar{s}_m est à $+0,0211$ écart-type de \bar{s}_0

6.1.5. $\ell = 100$ échantillons de taille $n = 100$

\bar{x}_m est à $-0,0061$ écart-type de \bar{x}_0

s_m est à $+0,0476$ écart-type de s_0 .

6.2. Essais en distribution exponentielle généralisée de types Goodrich et Frechet.

Les valeurs des paramètres d'échelle et de position des populations mères ont toujours été prises égales à 1 et 0, mais différentes valeurs du paramètre de forme δ ont été choisies.

\bar{x}_0 , \bar{s}_0 , \bar{x}_m et \bar{s}_m ont la même signification qu'en 6.1.

k_1 et k_2 étant les deux premiers cumulants de la population mère, γ_1 et γ_2 étant les coefficients d'asymétrie et d'aplatissement, pour ℓ échantillons de taille n les variances et covariances de \bar{x}_0 et \bar{s}_0 sont (approchées par défaut)

$$\text{variance } \bar{x}_0 = \frac{1}{n \ell} \left(k_2 - k_1 \sqrt{k_2} \gamma_1 + \frac{k_1^2}{4} (\gamma_2 + 2) \right)$$

$$\text{variance } \bar{s}_0 = \frac{1}{4 n \ell} (\gamma_2 + 2)$$

$$\text{covariance } (\bar{x}_0, \bar{s}_0) = \frac{1}{4 n \ell} (2 \gamma_1 \sqrt{k_2} - k_1 (\gamma_2 + 2)).$$

6.2.1. Paramètre de forme, connu à l'avance, $\delta = 0,28$

$$\text{d'où } \gamma_1 = 0,00744 \quad \gamma_2 = -0,28459 \quad \text{coef. de cor } (\bar{x}_0, s_0) \rho = -0,90298$$

$\ell = 100$ échantillons de taille $n = 10$

\bar{x}_m est à $-0,0472$ écart-type de \bar{x}_0

s_m est à $-0,0703$ écart-type de \bar{s}_0

$\ell = 100$ échantillons de taille $n = 25$

\bar{x}_m est à $-0,5889$ écart-type de \bar{x}_0

\bar{s}_m est à $+0,4973$ écart-type de \bar{s}_0

6.2.2. Paramètre de forme, connu à l'avance, $\delta = 0,175$

$$\text{d'où } \gamma_1 = -0,34279 \quad \gamma_2 = -0,00915 \quad \rho = 0,9677$$

$\ell = 100$ échantillons de taille $n = 16$

\bar{x}_m est à $-0,1532$ écart-type de \bar{x}_0

\bar{s}_m est à $+0,1213$ écart-type de \bar{s}_0

$\ell = 100$ échantillons de taille $n = 63$

\bar{x}_m est à $-0,3766$ écart-type de \bar{x}_0

\bar{s}_m est à $+0,3542$ écart-type de \bar{s}_0

6.2.3. Paramètre de forme connu à l'avance $\delta = 0,44$

$$\text{d'où } \gamma_1 = 0,46907 \quad \gamma_2 = -0,00044 \quad \rho = -0,782$$

$\ell = 100$ échantillons de taille $n = 16$

\bar{x}_m est à $-1,3031$ écart-type de \bar{x}_0

\bar{s}_m est à $+0,9312$ écart-type de \bar{s}_0

$\ell = 100$ échantillons de taille $n = 63$

\bar{x}_m est à $-1,5839$ écart-type de \bar{x}_0

\bar{s}_m est à $+1,2591$ écart-type de \bar{s}_0

6.2.4. Paramètre de forme connu à l'avance $\delta = 0,90$

$$\text{d'où } \gamma_1 = 1,70804 \quad \gamma_2 = 4,21409 \quad \rho = -0,692$$

$\ell = 100$ échantillons de taille $n = 10$

\bar{x}_m est à $-1,7134$ écart-type de \bar{x}_0

\bar{s}_m est à $+1,1181$ écart-type de \bar{s}_0

$\ell = 100$ échantillons de taille $n = 25$

\bar{x}_m est à $-1,7285$ écart-type de \bar{x}_0

\bar{s}_m est à $+1,1980$ écart-type de \bar{s}_0

6.2.5. Paramètre de forme connu à l'avance $\delta = 0,10$

$$\text{d'où } \gamma_1 = 1,91034 \quad \gamma_2 = 7,97857 \quad \rho = -0,99725$$

$\ell = 100$ échantillons de taille $n = 10$

\bar{x}_m est à $+0,0598$ écart-type de \bar{x}_0

s_m est à $-0,0459$ écart-type de s_0

$\ell = 100$ échantillons de taille $n = 25$

\bar{x}_m est à $-0,1999$ écart-type de \bar{x}_0

s_m est à $+0,2150$ écart-type de s_0

6.3. REMARQUES SUR LES ESSAIS EN EXPONENTIELLE GÉNÉRALISÉE

6.3.1. Le coefficient de corrélation (approximatif) entre les paramètres de position et d'échelle est toujours négatif et assez élevé en valeur absolue, ce qui explique que lorsque la différence entre \bar{x}_0 et \bar{x}_m est assez grande, la différence entre \bar{s}_0 et \bar{s}_m est du même ordre de grandeur mais de signe contraire.

Si on élimine les déterminations correspondant aux plus fortes valeurs observées du test, les écarts diminuent sensiblement : par exemple pour $\delta = 0,90$ en ne conservant que $\ell = 90$ échantillons de taille $n = 25$

\bar{x}_m est à $- 1,3106$ écart-type de \bar{x}_0

\bar{s}_m est à $+ 0,9866$ écart-type de \bar{s}_0

6.3.2. Les cinq séries d'essai ci-dessus montrent que x_m est systématiquement inférieur à x_0 et s_m supérieur à s_0 . Les différences restent cependant faibles lorsque l'ordre du point de contact de la fonction de densité avec l'axe des abscisses à la borne (inférieure) de l'intervalle de définition de la variable est élevé, et les différences augmentent lorsque cet ordre diminue ; par exemple pour $\delta = 0,44$ dérivée première de la fonction de densité en $x = x_0$ nulle mais dérivée seconde infinie, pour $\delta = 0,90$ (courbe en toboggan) dérivées première et d'ordre supérieur : infinies.

6.4. VALEURS MOYENNES DU TEST DANS LES ESSAIS PRÉCÉDENTS

6.4.1. La valeur moyenne de la valeur du test (non minimisé) a été calculée pour chaque série de 100 échantillons de 6.2 et cette valeur moyenne comparée à celle de la distribution du test (cf. *infra*).

$\delta = 0,28$ échantillons de taille $n = 10$
valeur moyenne du test 2,63783 à $+ 0,428$ écart-type de E (distribution)

échantillons de taille $n = 25$
valeur moyenne du test 3,78963 à $+ 0,313$ écart-type

$\delta = 0,175$ échantillons de taille $n = 16$
valeur moyenne du test 3,53193 à $+ 1,019$ écart-type

échantillons de taille $n = 63$
valeur moyenne du test 4,95774 à $+ 0,599$ écart-type

$\delta = 0,44$ échantillons de taille $n = 16$
valeur moyenne du test 3,38031 à $+ 0,655$ écart-type

échantillons de taille $n = 63$
valeur moyenne du test 4,48132 à $- 0,128$ écart-type

$\delta = 0,90$ échantillons de taille $n = 10$
valeur moyenne du test 5,09695 à $+ 8,598$ écart-type

échantillons de taille $n = 25$
valeur moyenne du test 6,95814 à $+ 6,371$ écart-type

$\delta = -0,10$ échantillons de taille $n = 10$
valeur moyenne du test 2,68072 à $+ 0,571$ écart-type

échantillons de taille $n = 25$
valeur moyenne du test 4,03871 à $+ 0,789$ écart-type

6.4.2. La distribution de la valeur du test est fortement asymétrique, avec un coefficient de variation élevé. Il s'ensuit, car le nombre d'échantillons par série est assez petit, que les probabilités au dépassement des valeurs des écarts positifs donnés ci-dessus sont supérieures aux probabilités qu'on lirait sur une table de la distribution normale.

Hors le cas $\delta = 0,90$, aucun écart n'est significatif. Cependant la répétition du signe $+$ montre que la condition 1.2.4 n'est pas parfaitement tenue.

Le cas $\delta = 0,90$ confirme ce qui a été constaté en 6.3.2.

6.5. D'autres séries d'essais ont été effectuées : loi de distribution choisie pour la population mère et l'ajustement : exponentielle généralisée avec paramètre de position connu à l'avance, paramètres de forme et d'échelle calculés par la méthode du maximum de vraisemblance.

6.5.1. La valeur moyenne de la valeur du test (non minimisé) a été calculée pour chaque série de 100 échantillons et cette valeur moyenne comparée à celle de la distribution du test (cf. *infra*)

population mère $\delta = 0,28$	échantillons de taille $n = 10$
valeur moyenne du test 2,4663 à $-$	0,246 écart-type de E (distribution)
	échantillons de taille $n = 63$
valeur moyenne du test 4,1163 à $-$	1,752 écart-type
population mère $\delta = 0,175$	échantillons de taille $n = 10$
valeur moyenne du test 2,7816 à $+$	1,572 écart-type
	échantillons de taille $n = 63$
valeur moyenne du test 4,6239 à $+$	0,230 écart-type
population mère $\delta = 0,44$	échantillons de taille $n = 10$
valeur moyenne du test 2,6216 à $+$	0,649 écart-type
	échantillons de taille $n = 63$
valeur moyenne du test 4,6631 à $+$	0,383 écart-type
population mère $\delta = 0,90$	échantillons de taille $n = 10$
valeur moyenne du test 2,2555 à $-$	1,462 écart-type
	échantillons de taille $n = 63$
valeur moyenne du test 4,4638 à $-$	0,395 écart-type
population mère $\delta = -0,10$	échantillons de taille $n = 10$
valeur moyenne du test 2,5639 à $+$	0,317 écart-type
	échantillons de taille $n = 63$
valeur moyenne du test 4,7929 à $+$	0,890 écart-type.

6.5.2. Aucune des différences ci-dessus n'est significative et nous admettons que la distribution de la valeur du test est indépendante du rôle des deux paramètres ajustés (sauf lorsque la fonction de densité est en J, en U ou en toboggan).

7. DISTRIBUTION DU TEST

7.1. Les connaissances qu'on peut avoir, a priori, sur la distribution du test pour une taille d'échantillon connue sont très succinctes : la valeur du test, d'après sa construction, est toujours positive dans un domaine de variation limité inférieurement par une valeur peu différente de zéro et supérieurement par une valeur élevée par rapport à sa valeur moyenne.

7.2. Pour déterminer la distribution du test nous avons effectué des tirages au hasard dans une distribution normale. Chaque échantillon a été rangé, ses paramètres de position $x_0 = \frac{1}{n} \sum x_i$ et d'échelle $s_0 = \frac{1}{n} \sum (x_i - x_0)^2$ calculés, et la valeur du test calculée avec les variables réduites $\frac{x_i - x_0}{s_0}$.

Pour chaque série d'échantillons de même taille il a été calculé les moments naturels d'ordre 1 (moyenne) à 4 et les coefficients d'asymétrie et d'aplatissement de la distribution d'échantillonnage du test. Les valeurs du test ont été réparties en tranches de valeurs 0,5 entre 0 et 20.

7.3. Le nombre d'échantillons tirés dans chaque série n'a pas été très grand (compris entre 10 000 pour $n = 8$ et 2 000 pour $n = 200$) si bien que les variances des moments naturels calculés par séries sont très grandes.

Cependant le graphique du coefficient d'aplatissement en fonction du coefficient d'asymétrie conduit à essayer une loi gamma incomplète généralisée, car les répartitions en tranches montrent que la distribution est unimodale et en cloche.

7.4. A priori la valeur du paramètre de position a été choisie nulle, et la fonction de densité de cette distribution s'écrit :

$$\frac{1}{s^\alpha \Gamma(\gamma)} \left(\frac{x}{s}\right)^{\frac{\gamma}{\alpha} - 1} \exp\left(-\left[\frac{x}{s}\right]^{1/\alpha}\right) dx$$

α et γ étant des paramètres de forme
s étant un paramètre d'échelle.

Le mode de la distribution est $s(\gamma - \alpha)^\alpha$ et les moments par rapport à l'origine sont $s^k \frac{\Gamma(\gamma + k\alpha)}{\Gamma(\gamma)}$.

D'après les séries de tirages au hasard, nous avons choisi pour représenter le moment d'ordre 1 :

$$m_1 = 32,31 (\log(n + 1,5))^{0,1054} - 32,99 = s \frac{\Gamma(\gamma + \alpha)}{\Gamma(\gamma)} \quad (n \text{ de } 8 \text{ à } 200).$$

Pour chercher les déterminations de α et de γ , nous sommes partis d'une part, du fait que la variance de

$$\frac{m_2}{m_1^2} = \frac{\Gamma(\gamma + 2\alpha) \Gamma(\gamma)}{[\Gamma(\gamma + \alpha)]^2}$$

n'est pas exagérément grande et d'autre part en cherchant à minimiser les χ^2 calculés d'après les répartitions observées par tranches et les répartitions déduites de valeurs de γ et de α choisies.

7.5. En fin de compte il semble que la probabilité liée à la valeur du test puisse être représentée, pour des tailles d'échantillons de $n = 8$ à $n = 200$, avec une erreur possible de l'ordre de $\pm 0,0005$, par une distribution gamma incomplète de paramètre de position nul, de paramètres de forme :

$$\gamma = 400 \left(1 - (n - 1)^{0,875} \exp\left(-\sqrt{\frac{n}{2} - 1}\right) \right)$$

$$\alpha = 10 \left(1 - (2,4n - 3,8)^{0,7936} \exp\left(-\sqrt{n - 2}\right) \right)$$

et d'échelle : $s = m_1 \frac{\Gamma(\gamma)}{\Gamma(\gamma + \alpha)}$.

8. UTILISATION DU TEST, CONCLUSION

8.1. Tel qu'il a été bâti le test semblerait ne devoir être utilisé que lorsque les valeurs de deux paramètres ont été ajustées. Il est certain que la distribution de la valeur du test change avec le nombre des paramètres ajustés.

Mais la principale utilisation d'un test d'ajustement est d'abord de permettre le choix entre différentes formulations mathématiques, certaines demandant l'ajustement d'un seul paramètre, d'autres de deux paramètres, d'autres encore de trois ou plus. Il suffit dans ce cas de comparer les différentes valeurs du test : la plus faible est considérée comme celle du meilleur ajustement : moindre divergence entre l'échantillon observé et l'échantillon « idéal ».

```

GAMMA INCOMPLETE GENERALISEE    POUR XTEST3
AVEC GAMMA=400*(1-(N-1)**.875*EXP(-SQRT(N/2-1)))
      ALPHA=10*(1-(2.4*N-3.8)**.7936*EXP(-SQRT(N-2)))
      MOYENNE=32.31*(LOG(N+1.5))**.1054-32.99
PARAM ECH=MOYENNE*EXP(ALGAMA(GAMMA)-ALGAMA(GAMMA+ALPHA))
PARAM POS=ZERO

```

TAILLE	GAMMA	ALPHA	MOYENNE	VARIANCE	C VAR	C ASY	C APL
8	11.581	2.478	2.205	2.703	0.745	2.053	7.561
10	29.803	3.580	2.509	3.006	0.691	2.034	7.671
13	62.847	4.981	2.850	3.623	0.668	2.046	7.912
16	96.538	6.052	3.107	4.171	0.657	2.049	7.991
20	138.130	7.094	3.372	4.736	0.645	2.031	7.900
25	182.740	7.978	3.626	5.227	0.631	2.000	7.616
32	232.125	8.741	3.894	5.678	0.612	1.928	7.173
40	273.760	9.238	4.124	6.003	0.594	1.883	6.711
50	310.178	9.577	4.345	6.286	0.577	1.836	6.221
63	340.853	9.787	4.565	6.559	0.561	1.778	5.820
79	363.448	9.902	4.770	6.887	0.550	1.707	5.548
100	379.668	9.962	4.976	7.234	0.541	1.680	5.302
126	389.595	9.987	5.169	7.639	0.535	1.645	5.199
158	395.126	9.996	5.351	8.070	0.531	1.640	5.109
200	398.039	9.999	5.533	8.572	0.529	1.629	5.046

8.2. Dans cette dernière optique on peut utiliser les probabilités au dépassement de la distribution du test lorsqu'on a plusieurs échantillons de tailles différentes pour lesquels on suppose a priori que la même formulation mathématique de la loi de distribution est valable, mais avec des valeurs numériques différentes suivant les échantillons des paramètres d'ajustement. Parmi les diverses formulations essayées, la meilleure est probablement celle pour laquelle la somme des probabilités au dépassement est la plus forte.

8.3. CONCLUSION

Le test d'ajustement proposé peut être considéré comme non paramétrique (signification indépendante de la formulation mathématique de la loi de distribution continue choisie) lorsque la fonction de densité et sa dérivée première sont nulles aux bornes de l'intervalle de variation de la variate.

Sa valeur est quasi minimale lorsque les paramètres d'ajustement sont calculés par la méthode du maximum de vraisemblance.

Son mode de calcul donne plus de poids aux valeurs extrêmes (inférieures et supérieures) de l'échantillon rangé et permet de tester l'ajustement sur une moitié seulement (inférieure ou supérieure) de l'échantillon rangé. Mais il n'est utilisable que pour des échantillons de taille comprise entre 8 et 200.

Comme expliqué en annexe, on peut l'appliquer sur une moitié — inférieure ou supérieure — de l'échantillon.

9. ANNEXES

9.1. SOUS ROUTINE XTEST3 (UT, F, N) PIÈCE JOINTE

Calcule la valeur du test UT.

Fournir en entrées la taille N de l'échantillon et les N probabilités F déduites des valeurs observées, rangées en ordre croissant, par la formulation mathématique choisie et les valeurs numériques des paramètres ajustés.

9.2. FONCTION QTEST3 (UT, M) PIÈCE JOINTE

Calcule la probabilité au non dépassement de la valeur UT provenant de XTEST3 ou de XTESM3.

Fournir UT et la taille M de l'échantillon.

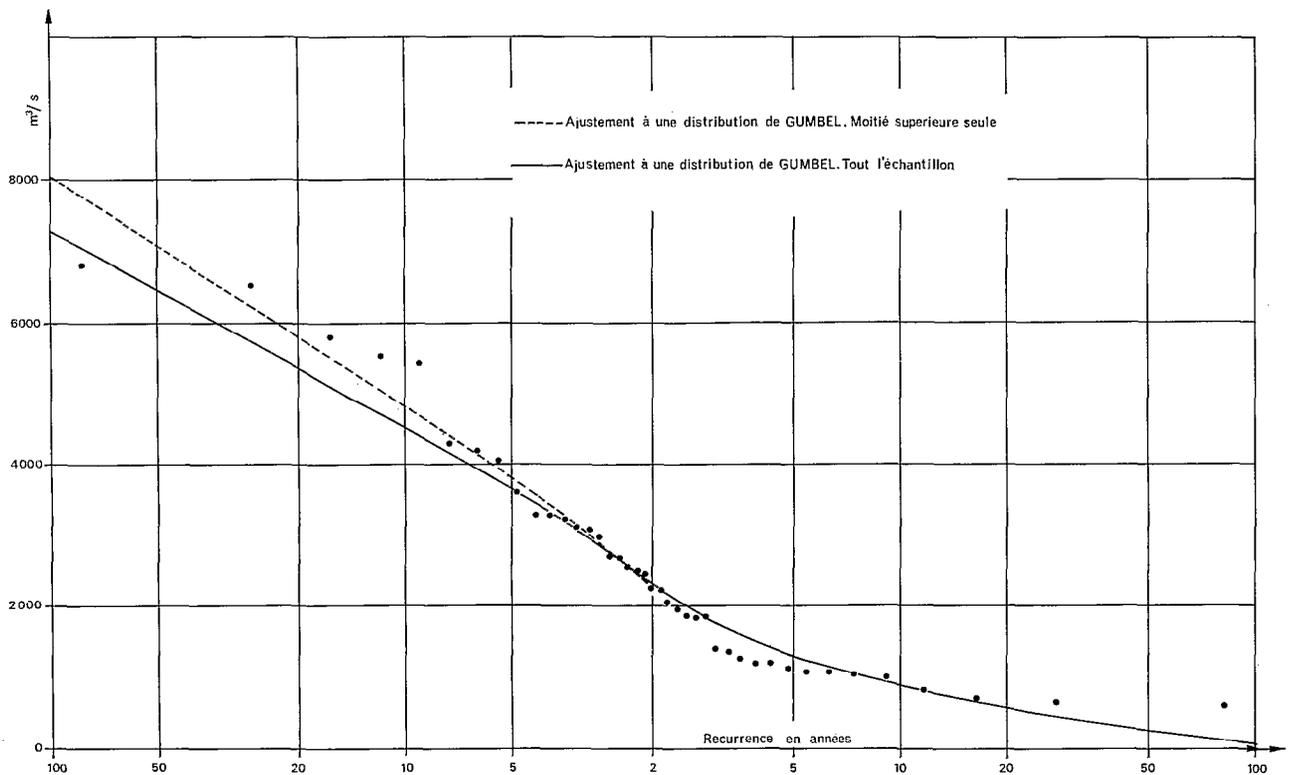


Fig. 7. — Débits instantanés (m³/s) maximaux d'années hydrologiques à Mjara (Ouergha, Maroc) 1932-33 à 1974-75 moins 1947-48 et 1948-49. Echantillon de 41 valeurs

9.3. SOUS ROUTINE XTESM3 (UT, F, N, KA) PIÈCE JOINTE

Calcule la valeur du test UT sur une moitié de l'échantillon.

Fournir en entrées la taille N de l'échantillon (total), les N probabilités F déduites des valeurs observées, rangées en ordre croissant, par la formulation mathématique choisie et les valeurs numériques des paramètres ajustés, et la valeur KA = 0 si on teste sur la moitié inférieure, ou KA = 1 si on teste sur la moitié supérieure.

9.4. EXEMPLE D'UTILISATION

Etude des débits instantanés maximaux d'années hydrologiques à la station de M'Jara (Ouergha, affluent du Sébou, Maroc) en distribution de Gumbel (fig. 7).

9.4.1. Le programme principal ci-joint entre les 41 valeurs X de l'échantillon rangées en ordre croissant, calcule par la sous routine PGUMMS les valeurs des paramètres d'ajustement X0 et S sur la moitié supérieure de l'échantillon [de X(21) à X(41) avec poids 1/2 pour X(21) et poids 1 pour les 20 autres valeurs], calcule les 41 valeurs F de la probabilité au non dépassement des valeurs X, X0 et S étant connus, calcule par la sous routine XTESM3 la valeur UT du test pour la moitié supérieure de l'échantillon (KA = 1) et par QTEST3 la probabilité au dépassement de la valeur UT.

9.4.2. Un autre programme, non fourni, mais dont les sorties ont été jointes à celles du précédent, calcule par la sous routine PGUMB les valeurs des paramètres d'ajustement X0' et S' sur la totalité de l'échantillon, calcule les 41 valeurs F' de la probabilité au non dépassement des valeurs X, calcule par la sous routine XTEST3 la valeur UT' du test pour la totalité de l'échantillon, la probabilité de UT' puis par la sous routine XTESM3 la valeur VT du test pour la moitié supérieure de l'échantillon (d'après les F') et la probabilité de VT.

Remarque : Test de χ^2 , 8 classes d'égaux probabilités : 5,244 probabilités (5 degrés de liberté) : 0,39
10 classes d'égaux probabilités : 7,049 probabilité (7 degrés de liberté) : 0,42.

```

RELEASE 2.0                XTST3                DATE = 78005                11/38/50

SUBROUTINE XTEST3(UT,F,N)
DIMENSION F(N),V(2,200),Z(202)
M=N+1
C=0.
DO 3 K=1,2
DO 2 I=1,N
IF(K.EQ.1.AND.I.GT.M/2)GO TO 3
G=2.*F(I)-1.
IF(K.EQ.1)G=2.*(I-.3)/(N+.4)-1.
V(K,I)=0.
1 T=FRF(V(K,I)/1.41421356)-G
V(K,I)=V(K,I)-T*EXP(V(K,I)*V(K,I)/2.)*.886226925
IF(4*INT(T*1.E5).NE.0.)GO TO 1
IF(K.EQ.2)GO TO 2
V(1,M-I)=-V(1,I)
C=C+V(1,I)*V(1,I)
2 CONTINUE
3 CONTINUE
C=SQRT((N+.204-1./1.8/(N+1.8)**1.5)*ALOG((N-1.8)/.2))/C/2.
Z(1)=0.
Z(N+2)=0.
DO 4 I=1,N
4 Z(I+1)=V(2,I)-V(1,I)*C
UT=0.
DO 5 I=2,M
VT=Z(I)
IF(7(I)*Z(I-1).LT.0.)VT=7(I)*Z(I)/(Z(I)-Z(I-1))
IF(7(I)*Z(I+1).LT.0.)VT=ABS(VT)+Z(I)*Z(I)/ABS(Z(I)-Z(I+1))
IF(7(I)*Z(I+1).GE.0.)VT=ABS(VT)+ABS(Z(I))
5 UT=UT+VT*VT
RETURN
END

```

RELEASE 2.0 QTFST3 DATE = 78005 17/12/59

```

FUNCTION QTEST3(UT,M)
  RFAI #4 A/S.755968/H/.6971471/.C/1.638484/.D/R.86007/.F/2.046555/.
  2UL*A.F/16.811243/.G*8,SS*A.SU*8
  QTEST3=.5
  GC=400.*(1.-(M-1.))**.875*EXP(-SGRT(M/2.-1.))
  AL=10.*(1.-EXP(.7976*ALOG(2.4*M-3.8)-SGRT(M-2.)))
  U=(UT/(32.31*(ALOG(M+1.5))**.1054-32.99)*EXP(DI GAMA(GC+AL)-ALGAMA(
  2GC)))*(1./AL)
  T=ALOG(GC)
  IF(U.LE.EXP((T*T-A)/(T-R)))QTEST3=1.
  IF(U.GE.EXP((C+T+SGRT(T*T+C*C+4.*C+2.*T*(C-2.*F)))/2.))QTEST3=0.
  IF(QTEST3.EQ.0..OR.QTEST3.EQ.1.)RETURN
  UL=NLOG(U*1.00)
  SI=SIGN(1.,2.*(U-F)*(U-F)*(U-F)+3.*(1.+F)*(1.+F))-GC*GC*GC)
  IF(SI.LT.0.)N=1.-GC+DEXP((UL+DSGRT(UL*(UL-4.*R)+4.*A))/2.)
  IF(SI.GT.0.)N=1.+GC-DEXP((UL*(UL-C)-D)/(UL-F))
  SS=0.00
  IF(N.LE.0)GO TO 2
  G=(1-SI)/2*GC+(1+SI)/2*AMAX1(1.F-7.GC-N)
  SU=1.00
  DO 1 I=1,N
  SU=SI*U/(G+I-1.)
  1 SS=SS+SU
  SS=SS*DEXP((G-1.00)*UL-DI GAMA(G)-U)
  2 QTEST3=(1.-SI)/2.+ST*SS
  RETURN
END

```

RELEASE 2.0 XTFSM3 DATE = 78005 17/12/59

```

SUBROUTINE XTFSM3(UT,F,N,KA)
  DIMENSION F(N),V(2,100),7(102)
  NN=(N+1)/2
  M=NN+1
  N1=1-KA+(N/2+1)*KA
  C=0.
  DO 2 K=1,2
  DO 2 I=1,NN
  G=2.*F(N1+I-1)-1.
  IF(K.EQ.1)G=2.*(N1+I-1.3)/(N+.4)-1.
  V(K,I)=0.
  1 T=FRF(V(K,I)/1.41421356)-G
  V(K,I)=V(K,I)-T*EXP(V(K,I)*V(K,I)/2.)*.886226925
  IF(AINT(T*1.E5).NF.0.)GO TO 1
  IF(K.EQ.1)C=C+V(1,I)*V(1,I)
  2 CONTINUE
  C=SQRT((N+(.204-1./1.8/(N+1.8))*.15)*ALOG((N-1.8)/.2))/C/2.)
  Z(1)=0.
  Z(NN+2)=0.
  DO 4 I=1,NN
  4 Z(I+1)=V(2,I)-V(1,I)*C
  UT=0.
  DO 5 I=2,M
  VT=Z(I)
  IF(Z(I)*Z(I-1).LT.0.)VT=Z(I)*Z(I)/(Z(I)-Z(I-1))
  IF(Z(I)*Z(I+1).LT.0.)VT=ABS(VT)+Z(I)*Z(I)/ABS(7(I)-Z(I+1))
  IF(Z(I)*Z(I+1).GE.0.)VT=ABS(VT)+ABS(Z(I))
  5 UT=UT+VT*VT*2.
  IF(N/2.LT.NN)UT=UT-Z((NN-1)*(1-KA)+2)*Z((NN-1)*(1-KA)+2)*4.
  RETURN
END

```

RELEASE 2.0

MAIN

DATE = 78005

17/12/89

```

1 FORMAT('1 DEBITS INSTANTANES M3/S MAXIMAUX D ANNEES HYDROLOGIQUES:
2 /' A MJARA (OUERGHAMAROC) 1932-3 A 1974-5 MOINS 1947-8 ET 48-9)
3 /' ECHANTILLON DE 41 VALEURS)
2 FORMAT('0 AJUSTEMENT A UNE DISTRIBUTION DE GUMMEL MOITIE SUPERIEURE
2 SEULE //) PARAMETRES DE POSITION: F10.3, ' D ECHELLE: F10.3/)
3 FORMAT('1 OBSERVE RECURRENCE ANS OBSERVE)
4 FORMAT(F7.0,F14.3,F11.0)
5 FORMAT('0 CALCULE RECURRENCE ANS CALCULE)
6 FORMAT('//) VALEUR TEST: F8.3, ' PROP AU DEPASSEMENT: F7.3)
REAL*4 P(9)/.01/.02/.05/.1/.2/.25/.3/.4/.5/
REAL*4 X(41)/600./665./717./817./1000./1040./1070./1080./1140./120
20./1200./1250./1350./1400./1830./1830./1850./1940./2050./2200./225
30./2440./2465./2540./2650./2700./2950./3040./3100./3210./3280./328
40./3600./4050./4180./4300./5440./5540./5800./6530./6800./
DIMENSION F(41)
N=41
WRITE(6,1)
WRITE(6,3)
DO 10 I=1,21
A=N/(I-.5)
10 WRITE(6,4)X(I),A,X(42-I)
CALL FGUMMS(X,S,XC,N)
WRITE(6,2)X0,S
WRITE(6,5)
DO 12 I=1,9
A=1./P(I)
B=X0-S*ALOG(-ALOG(P(I)))
C=X0-S*ALOG(-ALOG(1.-P(I)))
12 WRITE(6,4)B,A,C
DO 14 I=1,N
14 F(I)=EXP(-EXP(-(X(I)-XC)/S))
KA=1
CALL XTESM3(UT,F,N,KA)
A=QTST3(UT,N)
WRITE(6,6)UT,A
STOP
END

```

DEBITS INSTANTANES M³/S MAXIMAUX D ANNEES HYDROLOGIQUES
A MJANA (OUERGHAMAROC) 1932-3 A 1974-5 MOIS 1947-8 ET 48-9
ECHANTILLON DE 41 VALEURS

OBSERVE	RECURRENCE ANS	OBSERVE
600.	82.000	6800.
665.	27.333	6530.
717.	16.400	5800.
817.	11.714	5540.
1000.	9.111	5440.
1040.	7.455	4300.
1070.	6.308	4180.
1080.	5.467	4050.
1140.	4.824	3600.
1200.	4.316	3280.
1200.	3.905	3280.
1250.	3.565	3210.
1350.	3.280	3100.
1400.	3.037	3040.
1830.	2.828	2950.
1830.	2.645	2700.
1850.	2.485	2650.
1940.	2.343	2540.
2050.	2.216	2465.
2200.	2.103	2440.
2250.	2.000	2250.

AJUSTEMENT A UNE DISTRIBUTION DE GUMBEL. MOITIE SUPERIEURE SEULE

PARAMETRES DE POSITION 1749.831 D ECHELLE 1364.843

CALCULE	RECURRENCE ANS	CALCULE
-335.	100.000	8028.
-112.	50.000	7075.
252.	20.000	5804.
612.	10.000	4821.
1100.	5.000	3797.
1304.	4.000	3450.
1496.	3.333	3157.
1869.	2.500	2667.
2250.	2.000	2250.

VALEUR TEST 1.546 PROB AU DEPASSEMENT 0.931

AJUSTEMENT A UNE DISTRIBUTION DE GUMBEL. TOUT L ECHANTILLON

PARAMETRES DE POSITION 1868.540 D ECHELLE 1176.123

CALCULE	RECURRENCE ANS	CALCULE
72.	100.000	7279.
264.	50.000	6458.
578.	20.000	5362.
888.	10.000	4515.
1309.	5.000	3633.
1484.	4.000	3334.
1650.	3.333	3081.
1971.	2.500	2659.
2300.	2.000	2300.

VALEUR TEST 5.187 PROB AU DEPASSEMENT 0.251

MOITIE SUPERIEURE SEULE. VALEUR TEST 5.620 PROB 0.207