INTERVENTION DE M. JACQUES BERNIER

SUR L'APPLICATION DES DIVERSES LOIS LIMITES DES VALEURS EXTRÊMES AU PROBLÈME DES DÉBITS DE CRUE

I. — Introduction

L'étude des valeurs extrêmes d'un échantillon de n variables aléatoires n'est apparue qu'assez tardivement dans les travaux des statisticiens; pendant longtemps, la nature aléatoire même de ces variables a été mise en cause et ce n'est que vers les années 1925 à 1930 que des idées précises ont pu être dégagées sur ce sujet; le mérite en revient à d'éminents mathématiciens comme MM. FRÉCHET, FISHER et TIFFETT et surtout GUMBEL qui, par l'étude approfondie de la loi exponentielle qui porte son nom et par les nombreuses applications faites principalement en hydrologie statistique, a contribué le plus à la généralisation de cette théorie.

Dans la suite de cette note, nous allons faire quelques remarques sur la convergence vers les lois limites des valeurs extrêmes; cette question a soulevé, à maintes reprises, de nombreuses discussions.

Nous mettrons ensuite en évidence le lien qui existe entre la méthode d'estimation des crues basées sur la théorie des valeurs extrêmes et la méthode de M. GIBRAT basée sur la loi de Galton.

Nous étudierons enfin une des trois lois des valeurs extrêmes, dite loi de Fréchet, qui ne semble pas avoir été utilisée jusqu'à présent, mais dont certaines propriétés remarquables font qu'elle peut être appliquée avec succès dans certains cas, comme le montreront quelques exemples.

II. — Formes limites générales de la loi de la plus grande valeur d'un échantillon de n variables aléatoires indépendantes

Toutes les formes limites possibles de la fonction de répartition de la plus grande valeur peuvent être obtenues (cf. FISHER et TIPPETT) par la résolution de l'équation fonctionnelle suivante :

$$F^n(x) = F(a_n x + b_n) \qquad \text{(II, 1)}$$

Les solutions de (II, 1) sont de trois types différents :

— TYPE I (loi de Gumbel) : F (x) de la forme $e^{-e^{-x}}$; le champ de variation de la variable est $-\infty$, $+\infty$. La loi de la plus grande valeur d'un échantillon de grandeur n tiré de cette loi est de même forme que la loi initiale; celle-ci subit simplement une translation de module $\log n$: ceci résulte de (II, 1) où $a_n = 1$, $b_n = \log n$.

— TYPE II (loi de Fréchet) : F (x) de la forme $e^{-x^{-k}}$ (k positif); le champ de variation est 0, $+\infty$.

Pour cette loi : $a_n = n^{-1/k}$, $b_n = 0$; l'échelle de la loi est seule modifiée lorsque l'on prend la plus grande valeur : elle est multipliée par $n^{1/k}$ (facteur croissant avec n).

— TYPE III : F (x) de la forme $e^{-(-x)^k}$ (k positif) : le champ de variation est : $-\infty$, 0.

La formule (II, 1) est telle que $a_n = n^{+1/k}$ $b_n = 0$: l'échelle est multipliée par $n^{-1/k}$ (facteur décroissant avec n).

Les conditions qui permettent alors de choisir entre ces diverses formes reposent sur le comportement asymptotique de la loi de l'échantillon initial.

Soit $F_1(x)$ la fonction de répartition de cet échantillon et $f_1(x)$ sa densité. Lorsque F_1 tend vers 1 :

Si f_1 décroît de façon exponentielle : la limite est le type I.

Si x tendant vers l'infini, f_1 décroît comme x^{-k}, la limite est le type II.

Si x tendant vers zéro par valeurs négatives, f_1 décroît comme $|x|^k$; la limite est le type III.

Article published by SHF and available at http://www.shf-lhb.org or http://dx.doi.org/10.1051/lhb/1956049

III. — Remarques sur la rapidité de convergence des lois exponentielles vers la loi de Gumbel

FISHER et TIPPETT ont étudié la convergence de la plus grande valeur de n variables normales par l'intermédiaire des coefficients β_1 et β_2 de Pearson dont les expressions en fonction des moments centrés μ_i sont :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} \qquad \text{(III, 1)}$$

La convergence de ces paramètres vers les valeurs correspondantes de la loi de Gumbel est extrêmement lente, comme le montrent les chiffres du tableau (III, 2).

TABLEAU (III, 2)

n	β_1	β_2
2	0,019	3,062
5	0,092	3,202
10	0,168	3,331
100	0,429	3,765
500	0,570	4,003
1 000	0,618	4,088
Loi de Gumbel	1,299	5,400

Il nous semble cependant hasardeux de conclure, comme l'ont fait FISHER et TIPPETT, à la lenteur de convergence des lois de probabilités elles-mêmes. En effet, il arrive souvent en statistique mathématique que la rapidité de convergence des fonctions de répartition ou des densités de probabilité et la rapidité de convergence de paramètres attachés à ces fonctions soient extrêmement différentes.

Tel est le cas de la médiane d'une loi normale : la densité de la médiane converge rapidement vers la densité normale, alors que la variance asymptotique $\pi\sigma^2/2n$ est atteinte très lentement, comme l'a montré HOJO (1931). Ceci peut expliquer dans une certaine mesure la contradiction apparente qui existe entre les conclusions de FISHER et TIPPETT et celles de M. GUMBEL qui s'appuient sur les propriétés de la médiane et des densités de probabilités. Mais il semble plus plausible d'expliquer ces différences par le fait que l'étude de M. GUMBEL porte sur la partie centrale des distributions alors que les coefficients β_1 et β_2, faisant intervenir des moments d'ordre élevés, sont fortement influencés par les grandes valeurs et caractérisent ainsi la convergence des « queues de distribution ». Ceci nous a amenés à penser que les écarts relatifs entre les lois exactes et lois limites étaient plus considérables pour les grandes valeurs de la variable.

Pour étudier d'une façon générale cette convergence des parties extrêmes des distributions, nous avons utilisé la durée de retour $T=1/[1-f(x)]$ qui fait plus nettement apparaître les écarts et qui a le mérite d'avoir une signification concrète importante en hydrologie.

Rappelons ici le comportement asymptotique de différentes lois. Nous empruntons les formules qui lient asymptotiquement la variable aléatoire Q et la durée de retour T à une étude de M. LE CAM.

$$\begin{array}{ll} \text{Loi de Gauss} & Q \sim \sqrt{\log T} \\ \text{Loi de Galton-Gibrat} & Q \sim \exp\sqrt{\log T} \\ \text{Loi III de Pearson} & Q \sim \log T \\ \text{Loi de Laplace} & Q \sim \log T \\ \text{Loi de Gumbel} & Q \sim \log T \\ \text{Loi de Fréchet} & Q \sim T^{1/k} \end{array} \qquad \text{(III, 3)}$$

D'après ces formules, il semble que les lois de Gauss et de Galton, qui ont des comportements asymptotiques très différents de la loi de Gumbel, doivent tendre lentement vers cette loi. En revanche, les lois III de Pearson, comme toutes les lois à décroissance simplement exponentielle, atteindront beaucoup plus rapidement la limite. Pour vérifier ceci, nous avons construit quelques graphiques.

a) CAS DE LA LOI NORMALE : figure 1.

Nous avons placé en abscisse, sur une échelle logarithmique, les durées de retour T, et en ordonnées les valeurs prises par la variable aléatoire (ce qui permet de représenter, pour les fortes valeurs, la loi de Gumbel par une droite).

Les courbes des lois exactes de la plus grande de 100 et même de 1 000 variables normales s'écartent très nettement des droites représentant la loi limite de Gumbel. La convergence est donc ici très lente et les résultats numériques sur les β_1 et β_2 ne sont plus surprenants.

Pour approcher de façon plus précise la loi exacte, FISHER et TIPPETT ont introduit une « penultimate form » que l'on peut traduire par forme pré-asymptotique : c'est une loi de probabilité qui, pour des valeurs de n modérées, permet une meilleure approximation que la loi limite. Ils ont alors montré que la loi III citée plus haut était une telle forme pré-asymptotique pour la loi de la plus grande de n variables normales.

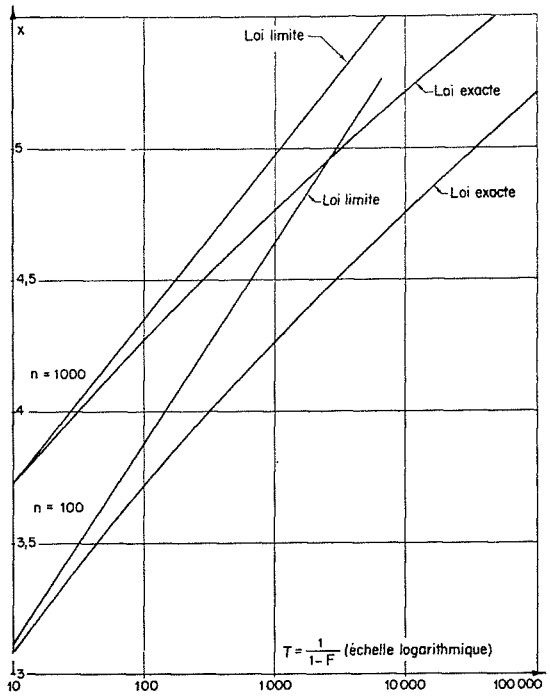

FIG. 1
Loi de la plus grande de $n = 100$ et $n = 1\,000$ variables normales.

b) CAS DE LA LOI DE LAPLACE, figure 2.

Nous avons tracé la loi limite et la loi exacte de la plus grande de 10 variables laplaciennes. Les courbes sont très proches l'une de l'autre. Pour cette loi, les coefficients β_1 et β_2 s'expriment d'ailleurs d'une façon relativement simple au moyen des premières dérivées de la fonction :

$$\psi(\gamma) = \frac{d}{d\gamma}\,[\log \Gamma(\gamma)]$$

$$\beta = \frac{[\psi''(n+1) - \psi''(1)]^2}{[\psi'(1) - \psi'(n+1)]^3} \qquad \text{(III, 4)}$$

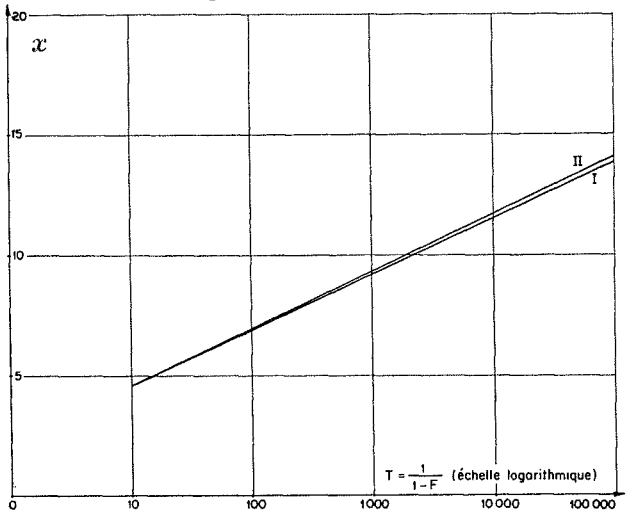

FIG. 2
Loi de la plus grande de 10 variables caplaciennes.
I. Loi exacte II. Loi limite

$$\beta_2 = 3 + \frac{\psi'''(1) - \psi'''(n+1)}{[\psi'(1) - \psi'(n+1)]^2} \qquad \text{(III, 5)}$$

Il est bien connu que les dérivées de la fonction ψ tendent rapidement vers zéro lorsque γ augmente, ce qui permet de voir que β_1 et β_2 convergent rapidement vers :

$$\frac{\psi''(1)^2}{\psi'(1)^3} \quad \text{et} \quad 3 + \frac{\psi'''(1)}{\psi'(1)^2}$$

qui sont les paramètres de même nom de la loi de Gumbel. Le calcul numérique effectué pour $n = 10$ donne $\beta_1 = 1{,}569$, $\beta_2 = 5{,}735$, alors que les valeurs limites sont, nous le rappelons, 1,299 et 5,400.

Ainsi la plus grande valeur d'un échantillon tiré d'une loi de Laplace (cas particulier d'une loi de Pearson) converge très rapidement vers la loi de Gumbel.

c) CAS DE LA LOI DE GALTON-GIBRAT : figure 3.

Le cas de cette loi est très intéressant car, comme l'a montré M. GIBRAT, la fonction de Gal-

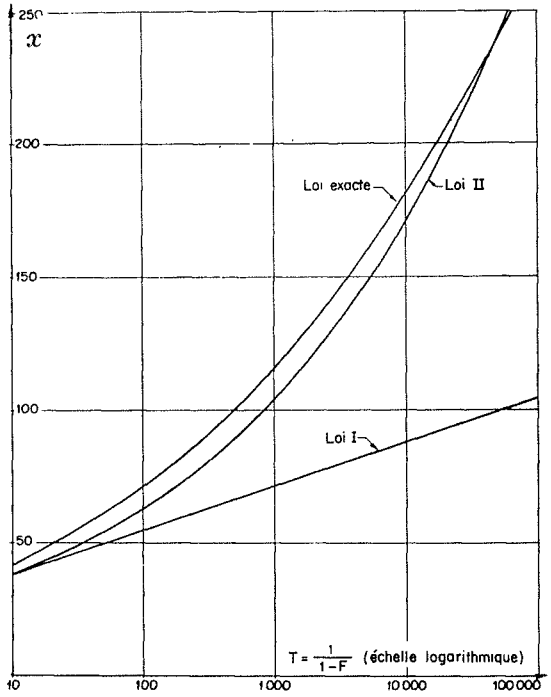

FIG. 3
Loi de la plus grande de 1 000 variables galtoniennes.
Loi I : $F = e^{-e^{-x}}$, Loi II : $F = e^{-x^{-k}}$

ton fournit en général un bon ajustement de la courbe des débits journaliers classés par ordre de grandeur. L'hypothèse de décroissance galtonienne semble donc tout à fait plausible pour la « loi » des débits journaliers. On peut démontrer que, dans ce cas, la loi limite de la plus grande valeur est encore la loi de Gambel. Ceci met en évidence le lien mathématique existant entre la méthode d'estimation des fortes crues au moyen

de la courbe de Galton ajustée aux débits classés et la méthode statistique utilisant la loi de Gumbel.

Toutefois, il convient de remarquer que la plus grande valeur de la loi de Galton converge vers la loi de Gumbel encore plus lentement que la plus grande valeur normale, comme le montrent les courbes tracées sur le graphique pour $n = 1\,000$. La loi de Galton tracée ici est telle que l'espérance mathématique et la variance du logarithme de la variable sont respectivement 0 et 1. Il faut donc tenir compte du fait que cette loi de Galton possède un très fort coefficient de variation.

De plus, la loi I sous-estime fortement la valeur correspondant à une durée de retour donnée surtout pour les très grandes valeurs; ceci peut présenter un grand danger pour l'estimation des débits de crue. On peut alors rechercher, comme l'ont fait Fisher et Tippett pour la loi normale, une forme pré-asymptotique qui s'adapte mieux pour des valeurs modérées de n à la loi exacte. Des calculs analogues à ceux de Fisher montrent que la loi de Fréchet est une telle forme. Nous avons placé sur le graphique cette loi (loi II) et la différence du comportement des trois lois est très significative. Il faut signaler que la loi de Fréchet étant à décroissance algébrique passe, pour les fortes valeurs, au-dessus de la courbe exacte.

IV. — Propriétés et ajustement de la loi de Fréchet

Ce que nous venons de voir au paragraphe précédent montre l'intérêt qu'il y a à considérer la loi de Fréchet. Nous avons déjà dit qu'elle n'avait pas été utilisée jusqu'à présent en hydrologie statistique; certains l'ont en effet rejetée parce qu'elle ne possède pas des moments d'ordre supérieur à k, étant donné sa décroissance algébrique. Nous allons cependant voir que cette propriété, entre autres, est particulièrement intéressante pour les applications à l'hydrologie statistique.

Revenons donc sur les principales propriétés de la loi de Fréchet :

a) Soit C_k la classe des fonctions de répartitions F_i telles que $x^k\ (1 - F_i)$ tend vers une constante ${}_k a_i$.

La loi de Fréchet est la loi limite de la plus grande de n variables aléatoires ayant chacune l'une quelconque des fonctions de répartitions F_i. Elle fait elle-même partie de la classe C_k, ce qui signifie que sa densité de probabilité décroît comme $1/x^k$ lorsque x tend vers $+\infty$; elle décroît donc plus lentement que toute fonction exponentielle. Ceci est intéressant, car on a reproché souvent à la loi de Gumbel sa décroissance trop rapide qui peut affecter des probabilités trop petites aux forts débits.

En effet, pour les grandes valeurs de la variable, nous pourrons écrire :

$$\log x = \frac{1}{k} \log T \qquad \text{(IV, 1)}$$

relation qui est à rapprocher de la formule analogue pour la loi de Gumbel :

$$x = \log T$$

L'hypothèse de décroissance algébrique des débits journaliers n'est pas dans certains cas à rejeter, mais elle n'est pas nécessaire pour l'application de la loi de Fréchet puisque nous avons vu au paragraphe précédent qu'elle se présente comme forme pré-asymptotique pour la plus grande de n variables galtoniennes. L'hypothèse de décroissance galtonienne peut donc suffire à justifier théoriquement cette loi.

b) La loi de Fréchet est stable : la plus grande valeur d'un échantillon de n variables tirées de cette loi à une loi de même forme : seule l'échelle est multipliée par $n^{1/k}$, comme le montrent les relations suivantes :

$$F^n = (e^{-x^{-k}})^n = e^{-(x/n^{1/k})^{-k}} \qquad \text{(IV, 2)}$$

c) La loi de Fréchet converge vers la loi de Gumbel lorsque k tend vers l'infini.

Pour étudier cette convergence, nous avons porté sur un graphique (fig. 4) en ordonnées la variable réduite y de la loi de Gumbel et en abscisses la variable $u = k\,(x - \tilde{x})$ où $\tilde{x}$ est le mode (valeur dominante) de la loi de Fréchet; la limite est représentée par la droite $y = u$.

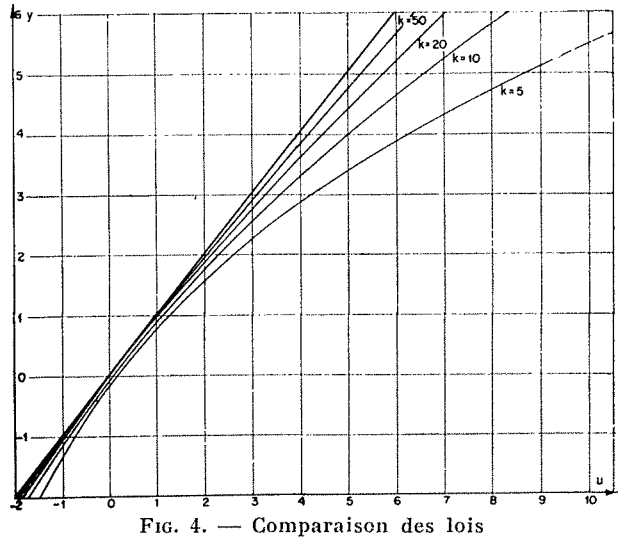

FIG. 4. — Comparaison des lois
I : $e^{-e^{-y}}$ et II : $e^{-x^{-k}}$ ou $u = k\,(x - \tilde{x})$

Nous avons ensuite tracé pour quelques valeurs de k la courbe correspondante à différentes lois de Fréchet. Le graphique montre clairement que cette loi converge assez rapidement vers la forme de Gumbel. Pour $k = 50$, les différences sont très petites; cependant, pour des k de l'ordre de 5 à 10, bien que les valeurs centrales soient très proches, les courbes s'éloignent considérablement pour les grandes valeurs de u.

d) On peut passer de la loi de Gumbel à la loi de Fréchet par la même transformation qui permet de passer de la loi normale à la loi de Galton, c'est-à-dire par une transformation logarithmique (il s'agit ici, comme dans la suite, de logarithmes naturels). En effet, on peut écrire :

$$e^{-x^{-k}} = e^{-e^{-k \log x}} = e^{-e^{-ku}} \quad \text{où} \quad u = \log x \qquad \text{(IV, 3)}$$

Cette importante propriété permet d'adapter à la loi de Fréchet les méthodes graphiques conçues par M. GUMBEL. Il suffit pour cela de remplacer l'échelle linéaire verticale du papier à probabilité par une échelle logarithmique.

Pour permettre l'ajustement de la loi aux observations, il faut introduire un second paramètre qui fixe l'échelle de la variable; la fonction de répartition peut alors se présenter sous les deux formes suivantes :

$$F = e^{-(ax)^{-k}} \qquad \text{(IV, 4)}$$

$$F = e^{-e^{-k(\log x - h)}} \qquad \text{(IV, 5)}$$

a et h sont liés par la relation $h = \log a$.

Pour ajuster la loi sous sa première forme, on peut utiliser la moyenne arithmétique m et la moyenne géométrique g :

$$m = \frac{\Sigma x_i}{n} \qquad \log g = \frac{\Sigma \log x_i}{n} \qquad \text{(IV, 6)}$$

la méthode des moments donne, en introduisant le nouveau paramètre $\lambda = m/g$:

$$e^{-\gamma/k}\,\Gamma\left(1 - \frac{1}{k}\right) = \lambda \qquad \text{(IV, 7)}$$

$$\frac{1}{a} = g\, e^{-\gamma/k}$$

la constante γ qui s'introduit dans ces formules est la constante d'Euler : $\gamma = 0{,}5772157$.

Pour la seconde forme, on utilisera la moyenne et l'écart-type des logarithmes :

$$M = \frac{\Sigma \log x_i}{n} \quad \text{et} \quad S = \sqrt{\frac{\Sigma (\log x_i - M)^2}{n - 1}} \qquad \text{(IV, 8)}$$

k et h seront alors déterminés par les relations :

$$M = h + \frac{\gamma}{k} \qquad \text{(IV, 9)}$$

$$S^2 = \frac{\pi^2}{6\,K^2}$$

C'est cette dernière méthode d'ajustement que nous avons employée pour les applications qui vont suivre.

V. — Applications numériques

Nous avons utilisé les méthodes présentées dans les paragraphes précédents pour étudier les crues de trois rivières différentes :

— Le Rhin à la station de Rheinfelden;

— Le Colorado à la station de Black Cañon;

— La Durance à la station de l'Archidiacre (Serre-Ponçon); pour cette dernière, la distinction entre débits de printemps et débits d'automne a été faite.

Le détail de l'ajustement des divers paramètres des lois de Fréchet et de Gumbel est donné dans le tableau (V, 1).

TABLEAU (V, I)

NOM DE LA STATION	LOI DE GUMBEL			LOI DE FRÉCHET			
	m	u	$\frac{1}{\alpha}$	M	S	K	h
RHEINFELDEN				7,824	0,2798	4,59	7,698
ARCHIDIACRE (crues de printemps)	312	259	92,21	5,683	0,3364	3,82	5,532
ARCHIDIACRE (crues d'automne)	202	154	81,94	5,167	0,5547	2,31	4,918

a) Le Rhin à Rheinfelden est particulièrement intéressant, car cette station possède une des plus longues séries connues de débits maxima annuels mesurés avec une bonne précision.

L'ajustement par la loi de Fréchet (fig. 5) est

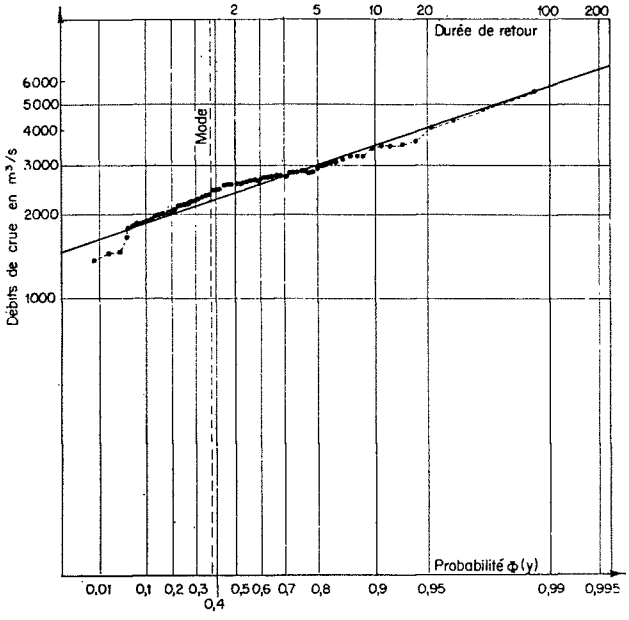

FIG. 5
Débits maxima annuels du Rhin à Rheinfelden
(Loi de Fréchet).

excellent alors que la loi de Gumbel s'écarte très nettement de la courbe observée pour les fortes crues (cf, note de M. GUMBEL).

b) Le Colorado à Black Cañon présente une série d'observations de longueur moyenne (52).

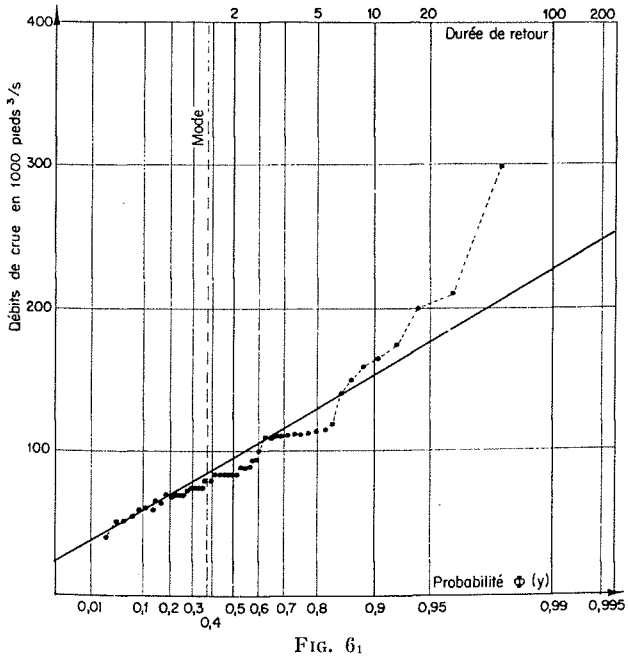

FIG. 6_1

Nous n'avons fait pour cette station qu'un ajustement graphique, mais la confrontation des deux ajustements (fig. 6_1 et 6_2) est significative. La loi de Fréchet est ici manifestement supérieure à la loi de Gumbel.

c) La Durance à l'Archidiacre ne présente qu'une courte série d'observations (40) et il est donc très difficile de tirer des conclusions précises de l'ajustement à ces 40 débits; quoi qu'il en soit, il semble que la loi de Fréchet soit préférable à la loi de Gumbel pour les crues de printemps (fig. 7_1 et 7_2) et la loi de Gumbel préférable à la loi de Fréchet pour les crues d'automne (fig. 8_1 et 8_2).

Mais, d'autre part, nous connaissons deux crues exceptionnelles survenues au siècle dernier, soit :

Une crue de printemps, 30 mai 1856 2 000 m³/s

Une crue d'automne, 2 novembre 1843 1 500 m³/s

Or, la loi de Gumbel affecte à la crue de 2 000 m³/s une durée de retour de 10^8 années; ce chiffre ne peut évidemment pas être pris en considération et le résultat de la loi de Fréchet (environ 2 000 ans) semble plus raisonnable.

De même la loi de Gumbel affecte à la crue d'automne de 1 500 m³/s une durée de retour de 10^7 ans, alors que cette crue est millénaire avec la loi de Fréchet.

Ceci permet de supposer que la loi de Fréchet doit être admise pour les crues de printemps et même pour les crues d'automne.

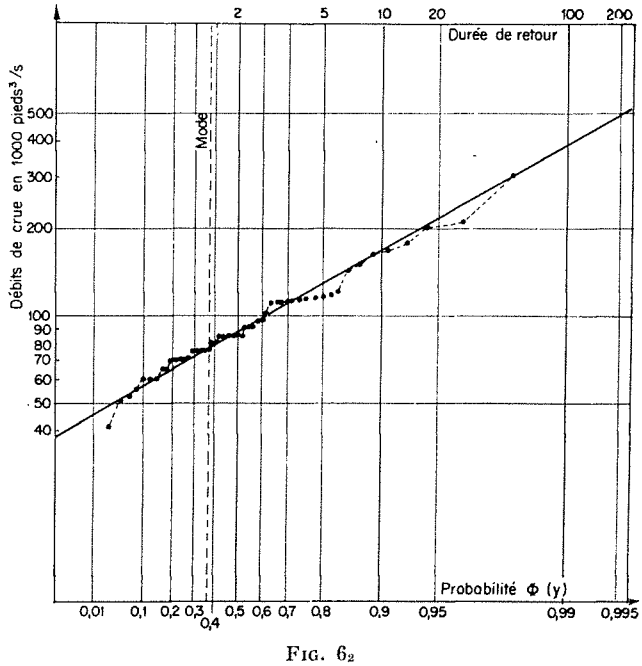

FIG. 6_2

Débits maxima annuels du Colorado à Black Cañon (1878-1929)
(Loi de Gumbel) (1878-1929) (Loi de Frechet)

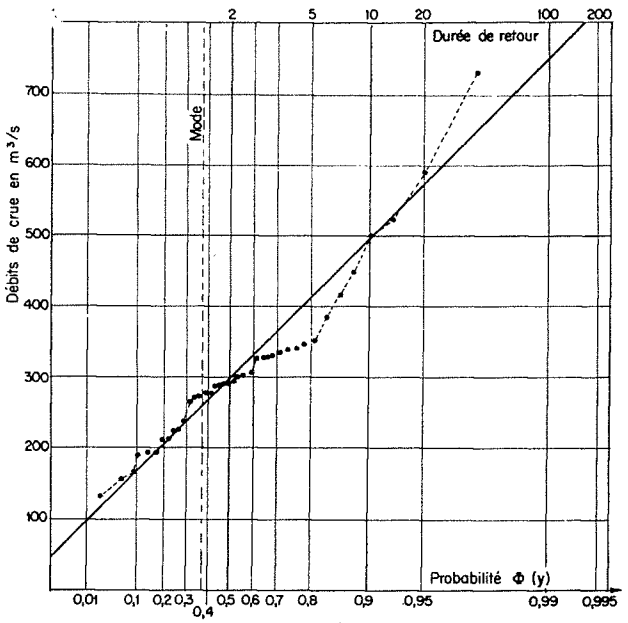

FIG. 7_1

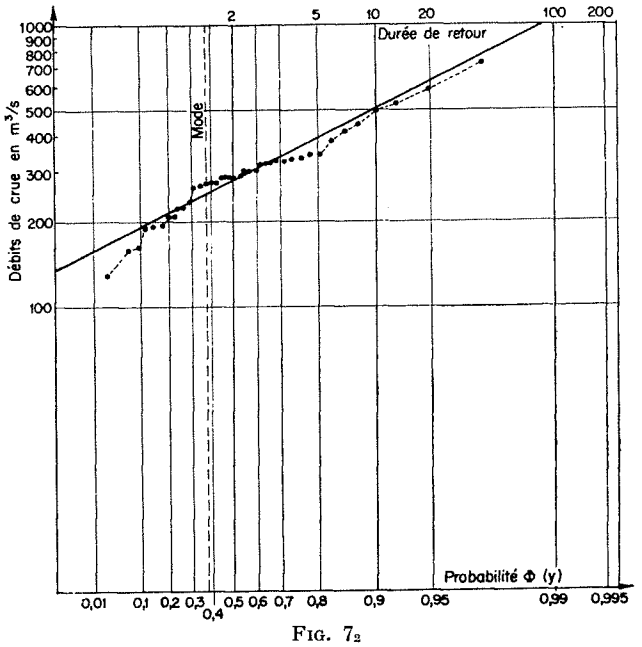

FIG. 7_2

Durance à l'Archidiacre. Crues de printemps 1916-1955

(Loi de Gumbel). (Loi de Fréchet).

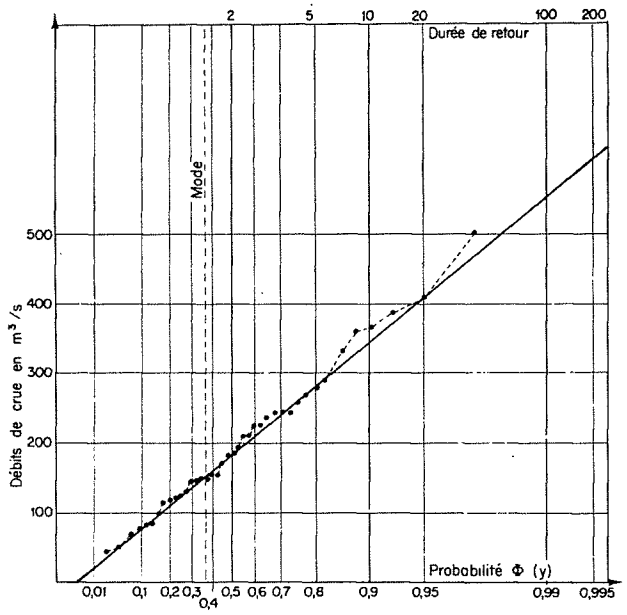

FIG. 8_1

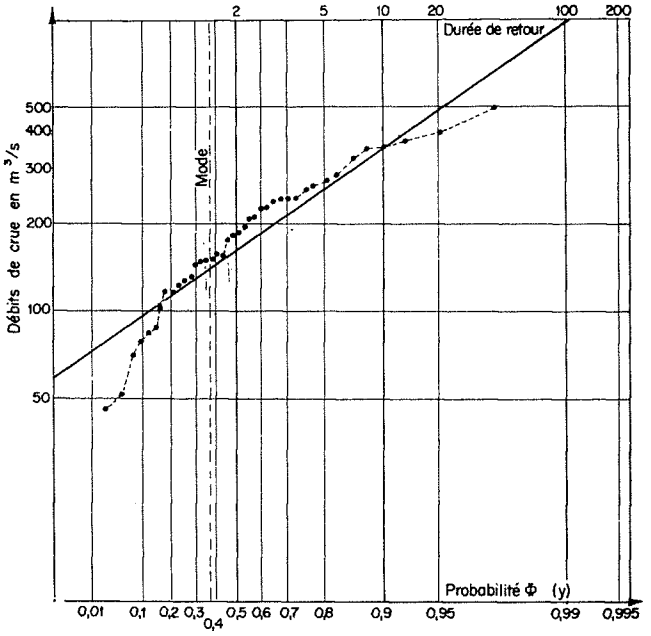

FIG. 8_2

Durance à l'Archidiacre. Crues d'automne 1916-1955

(Loi de Gumbel). (Loi de Fréchet).

VI. — Conclusion

Il est évident que l'emploi des lois de Fréchet et de Gumbel pour l'étude statistique des débits de crue est préférable à l'application d'autres lois (telle la loi de Galton-Gibrat) de par les justifications théoriques que l'on peut donner. Certains ont nié l'existence de telles justifications surtout en se référant à la lenteur de convergence vers la loi de Gumbel. Nous avons dit ce qu'il fallait en penser : compte tenu du fait que la loi de Fréchet peut s'appliquer lorsque la décroissance de la loi initiale est galtonienne, la convergence vers la loi de Gumbel n'est réelle-

ment lente que pour la loi normale et cette dernière loi ne peut prétendre représenter correctement une « loi » des débits journaliers. D'autre part, l'objection ne tient plus si on applique la loi de Fréchet puisque la convergence est alors très rapide.

Nous ne prétendrons pas que la loi de Fréchet est toujours préférable à la loi de Gumbel; il existe de nombreux exemples où cette dernière loi donne des résultats remarquables, mais il semble qu'à justifications théoriques égales, il y aura intérêt, dans certains cas, à utiliser la loi de Fréchet surtout en raison de la décroissance plus lente de cette dernière. Ceci pose le problème du choix *a priori* à faire entre les deux lois; en effet, il n'existe en général que de très courtes séries d'observations de crue et la comparaison directe des ajustements est malaisée. Mais nous avons vu que la convergence vers l'une ou l'autre de ces lois limites dépend de propriétés asymptotiques des distributions initiales; on peut donc espérer que l'étude des courbes de débits classés par ordre de grandeur permettra de dégager des critères *a priori* pour le choix des lois. Cette étude est actuellement en cours..

BIBLIOGRAPHIE

FISHER (R. A.) et TIPPETT (L. H. C.). — Limiting forms of the frequency distribution of the largest and smallest member of a sample (*Contribution to mathématical statistics*. Wiley N. Y., 1950).

FRECHET (M.). — Sur la loi de probabilité de l'écart maximum. *Annales de la Société polonaise de Mathématique*. Volume 6. Cracovie, 1927.

GUMBEL (E. J.). — Statistical theory of extreme values and some practical applications. *National Bureau of Standards*, Washington, 1954.

HOJO (T.). — Median, quartiles and interquartile distance of a sample from a normal population. *Biometrika*, 1931.

Importante société coloniale recherche un INGÉNIEUR HYDRAULICIEN ayant pratiqué hydrographie et topographie, pour diriger

MISSION D'ÉTUDES HYDROÉLECTRIQUES AU CONGO BELGE.

Ecrire avec curriculum vitae *détaillé, références et prétentions à :*

SOCIÉTÉ DES FORCES HYDRO-ÉLECTRIQUES DE L'EST DE LA COLONIE

47, rue de Trèves — BRUXELLES (Belgique)