

REVUE DE STATISTIQUE APPLIQUÉE

P. GUILLOT

Une extension des lois A de Halphen comprenant comme cas limite la loi de Galton-Gibrat

Revue de statistique appliquée, tome 12, n° 1 (1964), p. 63-73

http://www.numdam.org/item?id=RSA_1964__12_1_63_0

© Société française de statistique, 1964, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*
<http://www.numdam.org/>

UNE EXTENSION DES LOIS A DE HALPHEN COMPRENANT COMME CAS LIMITE LA LOI DE GALTON-GIBRAT

P. GUILLOT

Chef du Service Hydrométéorologique E.D.F., Grenoble

La transformation logarithmique s'impose quand on examine les distributions des débits journaliers des cours d'eau ; en effet, alors que la distribution des valeurs naturelles de ces débits, même pour des cours d'eau importants, est toujours très dissymétrique (forte densité en dessous de la médiane, faible densité et grande dispersion au-dessus), celle de leurs logarithmes est habituellement beaucoup plus proche de la symétrie.

Admettre que les logarithmes suivent une loi normale est alors une hypothèse commode, d'autant plus tentante que M. GIBRAT en a proposé une justification théorique par la "loi de l'effet proportionnel", selon laquelle les effets des événements aléatoires sur le débit seraient proportionnels au débit lui-même. Ceci n'est évidemment qu'un schéma - disons simplement qu'il est exact que les effets croissent avec le débit - mais il faut reconnaître que, pour beaucoup d'applications simples, la transformation logarithmique suffit et qu'il est commode d'examiner les distributions des débits journaliers sur un fond gradué gaussien-logarithmique, sans croire pour autant à la validité complète de la loi log-normale.

En effet, si on examine de plus près les distributions expérimentales, on s'aperçoit :

1/ que la symétrie des logarithmes n'est pas toujours admissible. Or, il n'existe aucun moyen satisfaisant de rendre la loi log-normale dissymétrique ; le déplacement de l'origine en prenant non plus $\log q$ mais $\log (q - q_0)$ est inacceptable, car d'une part il n'y a aucune raison d'admettre que le débit d'un cours d'eau possède une limite inférieure autre que 0, d'autre part, l'ajustement opéré par cet expédient n'est réellement pas satisfaisant : il le serait, à la rigueur, pour une variable aléatoire avec faible coefficient de variation qui ne prendrait jamais de valeurs voisines de q_0 ou de 0, mais ce n'est pas le cas des débits journaliers ni même des débits mensuels sur bien des cours d'eau.

2/ Même dans les cas où la symétrie des logarithmes est admissible, il apparaît, à l'expérience de nombreux cours d'eau (zone méditerranéenne exclue) que la densité de la loi log-normale ne décroît pas assez vite, aussi bien dans les valeurs fortes que dans les valeurs faibles.

Ceci rejoint le fait que la loi de l'effet proportionnel, même seulement la loi de l'effet croissant, première approximation admissible dans les valeurs courantes, a des limitations évidentes dans les valeurs extrêmes : sinon, un cours d'eau tari ne devrait plus jamais couler, aussi fort

qu'il pleuve ; sinon, une averse tombant en période de hautes eaux pourrait engendrer un volume d'écoulement plus important que le volume précipité.

Telles sont donc quelques unes des raisons qui ont conduit les hydrologues à chercher "autre chose".

M. HALPHEN présentait en 1941 une loi symétrique en logarithme qu'il appelait "loi harmonique" :

$$dF = \frac{1}{2 K_0''(2\alpha)} e^{-\alpha \left(\frac{x}{\mu} + \frac{\mu}{x}\right)} \cdot \frac{dx}{x}$$

Pour rendre la symétrie logarithmique plus évidente et l'expression plus maniable, nous posons : $Z = L \frac{x}{\mu}$

$$dF = \frac{1}{2 K_0''(2\alpha)} e^{-2\alpha \operatorname{ch} Z} dZ$$

$K_0''(u)$ est une fonction de BESSEL modifiée d'ordre 0, définie par :

$$K_0''(u) = \int_{t=0}^{t=\infty} e^{-u \operatorname{ch} t} dt$$

On trouve dans les divers ouvrages sur les fonctions de BESSEL, notamment celui de M. PETIAU, la table numérique de $K_0''(u)$ et ses développements limités :

- pour u petit ($\leq 0,1$) $K_0''(u) \approx -Lu + 0,116$
- pour u grand (> 5) $K_0''(u) \approx \left(\frac{\pi}{2u}\right)^{\frac{1}{2}} \cdot e^{-u}$

μ paramètre d'échelle est la médiane de la distribution ;

α est le paramètre de forme qui détermine la dispersion : nous donnons sur le graphique 1 ci-joint la valeur du paramètre α en fonction du rapport des déciles.

L'extension de la loi harmonique :

$$A_\gamma dF = \frac{1}{2 K_\gamma''(2\alpha)} e^{-\alpha \left(\frac{x}{\mu} + \frac{\mu}{x}\right)} \left(\frac{x}{\mu}\right)^{\gamma-1} \cdot d\left(\frac{x}{\mu}\right)$$

est appelée famille des lois A de HALPHEN. Ces lois ont été étudiées et tabulées à ELECTRICITE DE FRANCE par M. HALPHEN et ses collaborateurs. Le signe de γ donne le sens de la dissymétrie logarithmique. Ici μ n'est pas la médiane, le rapport de la médiane à μ variant avec α et γ .

Là encore, l'expression est plus maniable si on effectue la transformation logarithmique.

$$Z = L \frac{x}{\mu}$$

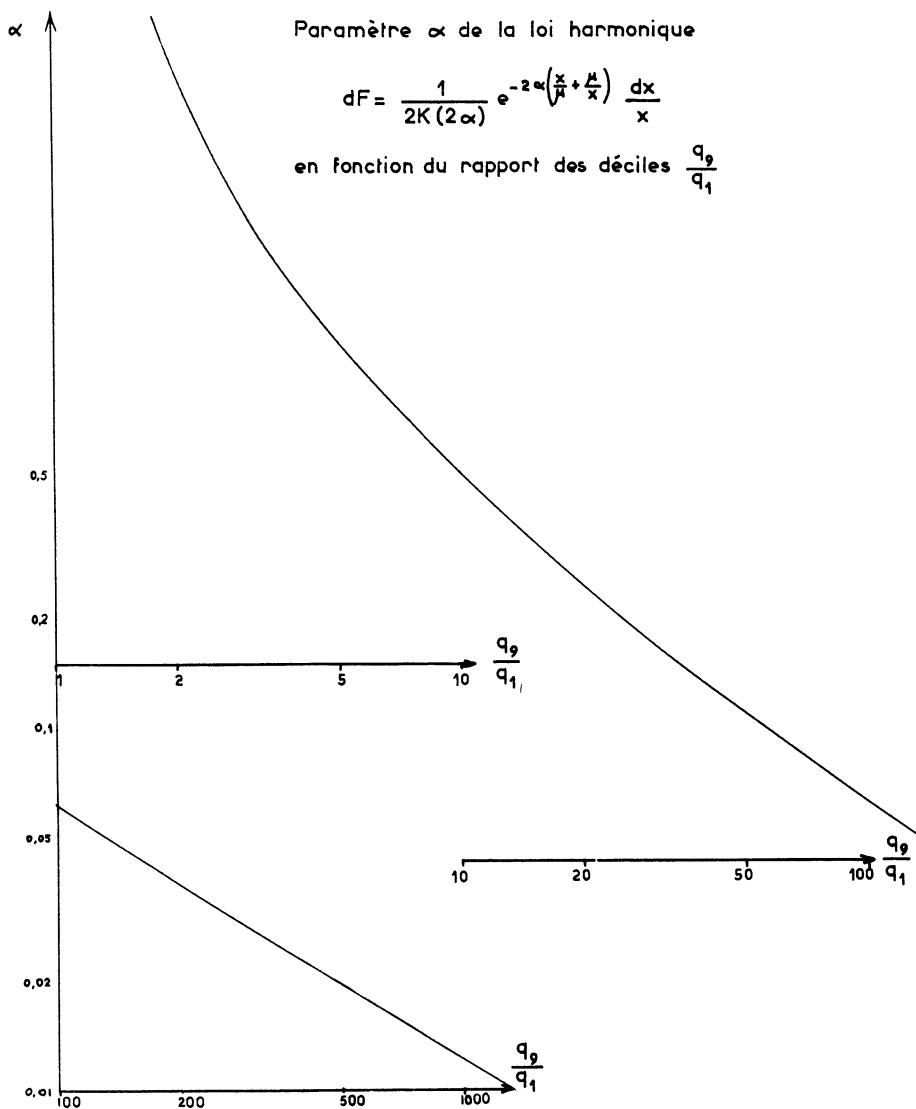


Fig. 1

$$A_\gamma \quad dF = \frac{1}{2 K_\gamma(2\alpha)} e^{-2\alpha hz + \gamma z} \cdot dz$$

$K_\gamma(2\alpha)$ est la fonction de BESSEL-HANKEL d'ordre γ :

$$K_\gamma(u) = \frac{1}{2} \int_{-\infty}^{+\infty} e^{-u \cosh t + \gamma t} \cdot dt$$

LOIS A DE HALPHEN

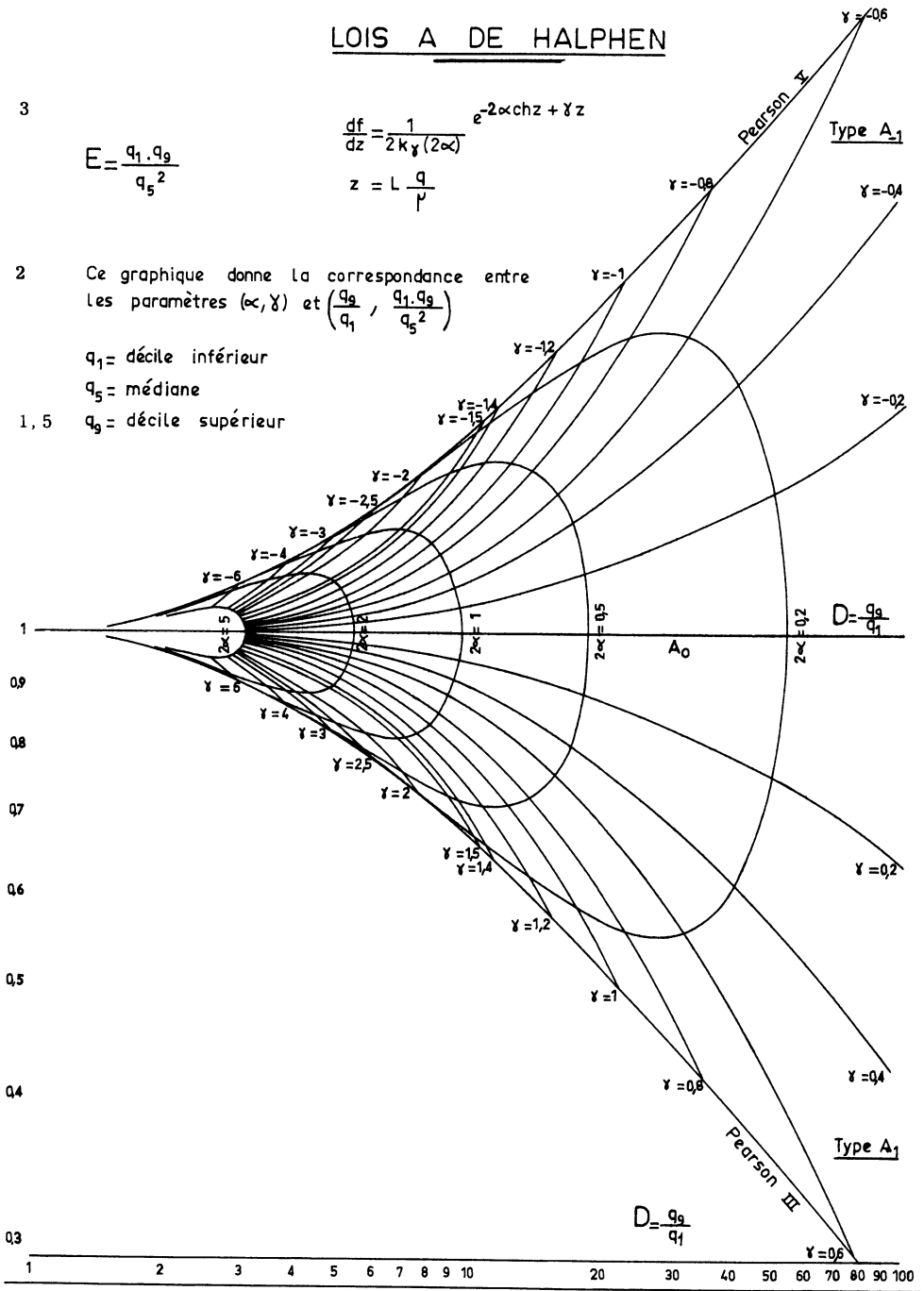


Fig. 2

Nous avons ébauché le graphique, figure 2 ci-joint, qui permet le choix des paramètres de forme α , γ à partir des fonctions des déciles $\frac{q_9}{q_1}, \frac{q_1}{q_5}, \frac{q_9}{q_5}$. Le rapport de la médiane q_5 à μ se trouve dans les tables.

Le graphique n'est pas complet : pour répondre aux besoins de l'utilisateur, il devrait comporter, dans 4 ou 5 régions, des surcharges figurant les limites de confiance de la distribution d'échantillonnage des fonctions quantiliques $\frac{q_9}{q_1}, \frac{q_1}{q_5}, \frac{q_9}{q_5}$: Ces distributions pourraient être établies empiriquement à partir d'échantillons de 25 nombres au hasard.

Ce graphique a été construit à partir des seules tables encore disponibles des lois de HALPHEN, tables sommaires calculées il y a une dizaine d'années par le Service des Etudes et Recherches : les intersections des courbes $\alpha = \text{Cte}$ et $\gamma = \text{Cte}$ qui figurent sur ce graphique correspondent aux lois tabulées.

A partir de là, on peut songer à une autre extension de la loi A de HALPHEN :

$$A_{\frac{1}{\nu}} \quad dF = \frac{1}{2 K_{\gamma} (2 \alpha)} e^{-\left(2 \alpha \chi \frac{Z}{\nu} - \frac{Z}{\nu}\right)} d\left(\frac{Z}{\nu}\right)$$

ce qui revient simplement à élever x à la puissance $\frac{1}{\nu}$. Cette extension comprend en particulier la loi log-normale qu'on obtient pour $\frac{1}{\nu} = 0$.

Pour montrer que la forme limite est bien la loi log-normale, il suffit d'observer que lorsque ν croît, $\chi \frac{Z}{\nu}$ peut-être confondu avec $1 + \frac{Z^2}{2 \nu^2}$ sur un intervalle centré de plus en plus grand. Naturellement, pour que la largeur de la distribution reste finie, on est conduit à faire croître simultanément α ; plus précisément, le coefficient $\frac{\alpha}{\nu^2}$ du terme en Z^2 tendant vers une limite $\frac{1}{2 \sigma^2}$, $\frac{\nu}{\sqrt{2 \alpha}}$ tend vers σ . On vérifie dans ces conditions, en remplaçant l'infiniment petit $K_0 (2 \alpha)$ par sa partie principale, que l'expression inverse de la densité au centre $2 \nu K_0 (2 \alpha) e^{2 \alpha}$ tend bien vers $\sigma \sqrt{2 \pi}$.

Ce qu'il est plus intéressant de remarquer, d'un point de vue pratique, est qu'il n'est pas nécessaire de prendre un ν très grand pour que la loi $A_0^{\frac{1}{\nu}}$ soit pratiquement confondue avec la loi log-normale, dans les valeurs centrales, disons dans l'intervalle centré 5 à 95 %.

La figure 3 permet de comparer les densités de probabilité obtenue en faisant varier ν tout en maintenant le rapport des déciles $\frac{q_9}{q_1}$ égal à 25 (cette amplitude de variation est couramment observée en matière de débits journaliers) : on voit que la loi $A_0^{\frac{1}{2}}$ ($\nu = 2$) est très proche de la loi log-normale A_0^0 tandis que A_0^1 et A_0^2 sont sensiblement plus plates. La position relative des courbes de densité $A_0^{\frac{1}{\nu}}$ reste la même quand on fait

Comparaison des lois $A_0^{\frac{1}{\nu}}$
 ayant le rapport $\frac{x(F=0,90)}{x(F=0,10)}$ égal à 25 .

$$A_0^0 : \frac{dF}{dz} = \frac{1}{(1,25)\sqrt{2\pi}} e^{-\frac{1}{2} \frac{z^2}{(1,25)^2}}$$

$$A_0^{\frac{1}{4}} : \frac{dF}{dz} = \frac{1}{4K_0(2,2)} e^{-2,2 \operatorname{Ch} \frac{z}{2}}$$

$$A_0^1 : \frac{dF}{dz} = \frac{1}{2K_0(0,38)} e^{-0,38 \operatorname{Ch} z}$$

$$A_0^2 : \frac{dF}{dz} = \frac{1}{K_0(0,034)} e^{-0,034 \operatorname{Ch} 2z}$$

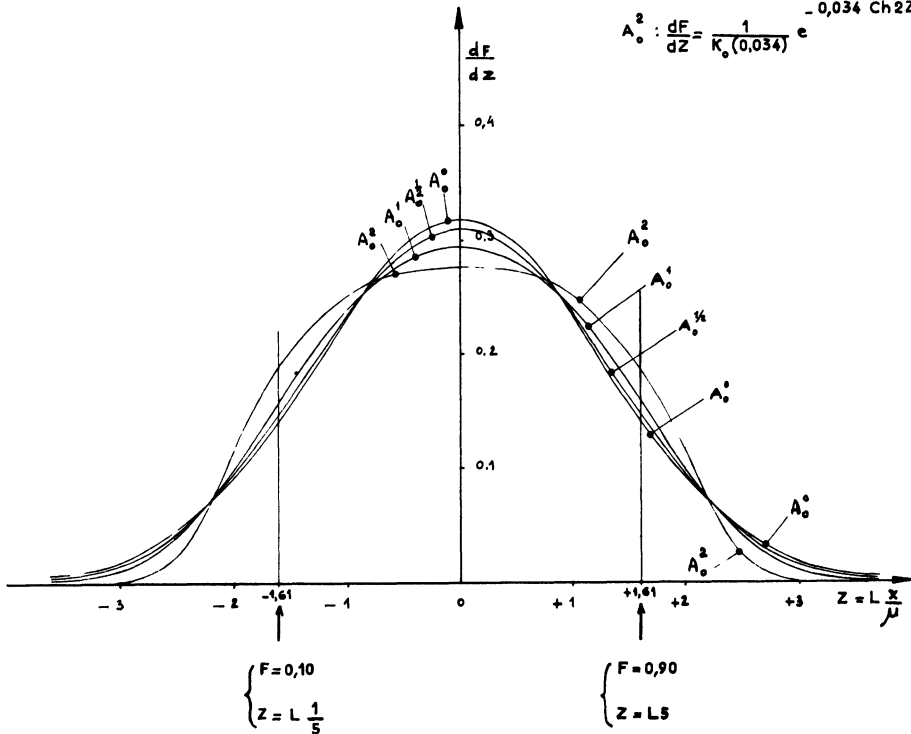


Fig. 3

varier la largeur de la distribution $\frac{q_0}{q_1}$; dans les valeurs centrales, les $A_0^{\frac{1}{\nu}}$ sont toujours plus plates que la log-normale A_0^0 de même "largeur", la platitude s'accroissant avec $\frac{1}{\nu}$.

Ainsi les 3 paramètres de forme ont chacun un rôle principal : α agit principalement sur la dispersion, γ sur la dissymétrie logarithmique, $\frac{1}{\nu}$ sur l'aplatissement.

Un paramètre de taille μ , et 3 paramètres de forme α , γ et ν , soit en tout 4 paramètres, c'est évidemment trop pour des échantillons hydrologiques de quelques dizaines d'observations : un paramètre de taille et 2 paramètres de forme constituent bien le maximum qu'on puisse raisonnablement envisager, et encore

S'il faut donc nous résigner à fixer ν , rien ne nous oblige à le prendre égal à 1 ; le schéma ci-dessous résume les conséquences du choix de ν sur la densité des valeurs éloignées de la zone centrale :

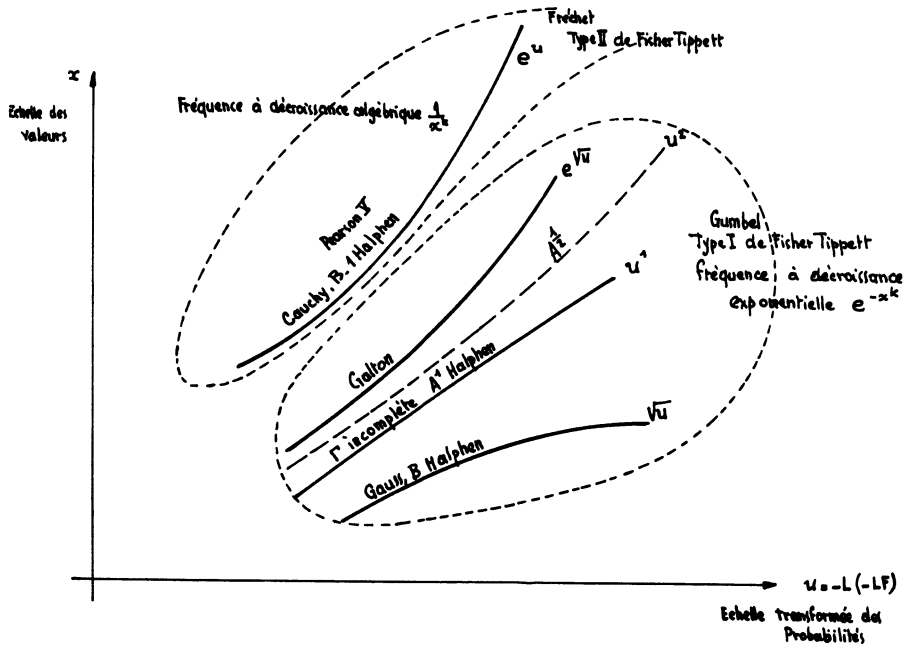


Fig. 4

Nous avons déjà signalé ailleurs que, en matière de débits journaliers, les valeurs extrêmes observées sur de nombreux cours d'eau (zone méditerranéenne exclue) se situent généralement au-dessus de la droite de GUMBEL ($\nu = 1$), mais en-dessous de l'hypothèse log-normale (que M. GIBRAT qualifie de prudente).

C'est pourquoi nous sommes tentés d'opter pour $\nu = 2$ qui nous donne :

- dans les valeurs courantes, une forme pratiquement identique à la log-normale quand il y a symétrie logarithmique,
- la possibilité d'introduire une dissymétrie logarithmique,
- un comportement raisonnable dans les valeurs extrêmes.

Il nous reste à souhaiter que la nouvelle tabulation des lois de HALPHEN, dont l'exécution par une calculatrice électronique est à l'étude, aboutisse à un outil complet et commode pour l'utilisateur.

REFERENCES BIBLIOGRAPHIQUES

- R. GIBRAT - Aménagement hydroélectrique des cours d'eau : statistique mathématique et probabilités - Revue Générale de l'Hydraulique - Septembre Octobre 1936
- R. GIBRAT - Débit maximum de crue et calcul des probabilités - C. rendus du quatrième Congrès des Grands Barrages - NEW DELHI 1951 - pp. 644-660
- Et. HALPHEN - Sur un nouveau type de courbe de fréquence - Compte rendu de l'Académie des Sciences - 10 Novembre 1941
- G. MORLAT - Les lois de probabilité de HALPHEN - Revue de Statistique Appliquée (Institut de Statistique de l'Université de PARIS) 1956 - n° 3
- G. PETIAU - La théorie des fonctions de BESSEL : C.N.R.S. - Institut Henri Poincaré - PARIS 1955
- P. GUILLOT - La probabilité du débit maximum annuel et ses relations avec la loi de distribution des débits journaliers - Congrès A.I.R.H. LONDRES - Septembre 1963

DISCUSSION

M. le Président remercie M. GUILLOT d'avoir clairement posé, à nouveau, le problème de l'adéquation des lois de distribution classiques et perfectionné celles qui avaient été proposées par M. HALPHEN ; ce dernier fut l'un des collaborateurs de la Société Hydrotechnique de France et l'animateur d'une brillante équipe de jeunes statisticiens intéressés à l'hydrologie.

La tabulation des fonctions d'Halphen a été commencée, il y a quelques années, par la Direction des Etudes et Recherches de l'E.D.F. ; elle vient d'être reprise en utilisant les facilités qu'offrent les ordinateurs électroniques.

Ayant trouvé très intéressante la remarque faite par M. GUILLOT à savoir que pour la plupart des rivières françaises qu'il avait étudiées les deux ou trois points les plus hauts de la courbe des crues annuelles classées se situaient systématiquement au-dessus de la droite d'ajustement de Gumbel et au-dessous de celle de Frechet, M. NORMAND suggère que ce critère empirique soit plus développé. La S.H.F. pourrait centraliser les ajustements de crues annuelles classées opérés par les divers groupes d'étude (E.D.F., O.R.S.T.O.M., Ponts et Chaussées, Génie Rural, Sociétés de Recherches Hydrauliques...) et animer un travail de synthèse qui permettrait de définir, en fonction des caractères climatiques et morphologiques des diverses régions, les lois statistiques les mieux adaptées en moyenne.

M. le Président dit que la vérification de l'adéquation des lois statistiques classiques à chaque cas particulier, est toujours assez délicate, du fait que l'on ne possède généralement que des données d'observation portant sur une faible durée (50 à 100 ans) alors que l'on recherche par extrapolation des valeurs correspondant à des temps de récurrence de 1 000 à 10 000 ans.

M. RODIER indique qu'il est assez fréquent que la loi log-normale qui, par ailleurs rend de grands services aux hydrologues, s'ajuste assez bien avec la distribution expérimentale des débits observés jusqu'au dernier décile et un peu au-delà mais ne convienne plus du tout pour les fréquences nettement plus faibles, Ceci n'empêche pas du tout 6 sur 10 des hydrologues (soyons optimistes) d'ajuster une relation de Galton sur les 20 ou 30 points expérimentaux et de l'extra-

poler hardiment vers la crue millénaire. M. GUILLOT a eu un grand mérite de rechercher une série de relations s'ajustant mieux avant le premier décile et au-delà du dernier. Bien sûr, on ne dispose que de 40 points en général pour voir si cela s'ajuste mieux mais le fait de trouver ce résultat pour une vingtaine de stations est déjà convaincant.

Quant à la recherche d'une valeur ν constante, M. RODIER précise ce qui suit : la loi de Gumbel donne en France des valeurs trop faibles. Nous avons trouvé ailleurs des régimes pour lesquels elle donnait des valeurs trop fortes. Par ailleurs, comme M. GUILLOT, M. RODIER a étudié d'autres régimes pour lesquels, même la loi de Galton donnait des résultats trop faibles. Il est possible que pour la majeure partie de la France il soit judicieux de prendre ν (nu) = 2 mais pour d'autres pays il faudrait admettre une valeur de ν (nu) différente.

M. SNEYERS demande :

1/ si les lois présentées peuvent attribuer une probabilité non nulle à une valeur de "x" nulle.

2/ s'il est possible d'ajuster de telles courbes à l'aide de la méthode de vraisemblance maximale.

D'autre part, M. SNEYERS indique, à propos de la remarque de M. RODIER, que l'appréciation de la validité de l'ajustement d'une loi donnée à un ensemble de rivières appartenant à un même régime risque de présenter des difficultés en raison du fait que le comportement des rivières n'est vraisemblablement pas indépendant.

M. GUILLOT répond négativement à la première question de M. SNEYERS car la probabilité d'avoir zéro est inférieure à toute quantité, si petite soit-elle, fixée d'avance ; M. GUILLOT répond affirmativement à la seconde question et dit qu'un des auditeurs a fait cet ajustement par la méthode du maximum de vraisemblance.

M. FERRY indique que, pour le statisticien pur, l'ajustement par les déciles choisi par M. GUILLOT n'est qu'une commodité de calcul, car les déciles se déterminent plus rapidement et facilement que les moments, mais au risque d'une perte d'information. Or, on peut se demander si, dans la réalité physique, ce n'est pas le contraire qui est vrai.

En effet, le statisticien commence par admettre, comme base de la plupart de ses raisonnements, la validité absolue d'une certaine forme mathématique de loi de probabilité, dont seule la valeur numérique des paramètres serait inconnue. Toutes les observations effectuées sont donc considérées par lui comme tirées d'une population obéissant rigoureusement à cette loi.

Si maintenant, nous examinons la réalité concrète, nous devons nous rappeler que les lois utilisées ont une justification purement empirique. Si même il existe certaines considérations théoriques, celles-ci ne sont jamais valables que comme des approximations à l'intérieur de limites que seule l'expérience peut déterminer.

Or, les données dont on dispose en hydrologie permettent en général de contrôler la validité pratique de tel ou tel type de loi à l'intérieur de limites correspondant à peu près à l'intervalle interdécile ; elles ne permettent pas d'étendre cette validité jusqu'aux centiles et encore moins au-delà.

Lorsqu'on utilise les moments pour le calcul des paramètres d'ajustement, on donne un poids considérable aux valeurs extrêmes pour lesquelles la validité de la loi admise est précisément la plus douteuse. (Ceci est particulièrement vrai lorsqu'on travaille sur les logarithmes d'une grandeur physique, pour les petites valeurs de cette grandeur).

Lorsqu'au contraire on utilise des quantiles pour l'ajustement, ils sont raisonnablement adaptés aux échantillons dont on dispose et ne risquent pas de faire commettre des erreurs d'adéquation à l'intérieur de l'intervalle interquantile utilisé.

La principale difficulté d'application de la statistique est de passer du domaine idéal et théorique du professeur de statistique mathématique au domaine réel et concret de l'art de l'ingénieur, et de savoir reconnaître les limites au-delà desquelles les méthodes de la théorie risquent de devenir fallacieuses.

M. LARRAS demande à M. GUILLOT s'il a pu dégager certaines impressions sur le type de loi le plus adéquat pour les régions méditerranéennes, de même qu'il en a dégagé pour les autres régions, (Gumbel, faible, Galton-Gibrat, fort).

M. GUILLOT répond que c'est assez difficile car, les stations de jaugeage étant souvent emportées par les crues, sur les cours d'eau méditerranéens, ceux-ci n'ont pas de séries de mesures de débits très longues.

D'après les estimations qui ont été faites se seraient plutôt des fréquences à décroissance algébrique qui conviendraient, genre hypothèse Fréchet (graphique 4).

Tout ce que l'on sait, en ce qui concerne les crues de la région méditerranéenne, c'est que cela dépasse vraiment les ordres de grandeurs les plus ahurissants, dans certains cas.

En conclusion, M. le Président propose que l'on tente de rechercher pour rechercher pour chaque région ou groupe de cours d'eau, hydrologiquement homogènes, le type de loi de distribution qui paraît le mieux adapté aux données disponibles. Sous certaines réserves évidentes, l'utilisation, en quelque sorte normalisée, de ce type de loi permettrait d'obtenir pour tous les problèmes pratiques (dimensionnement des évacuateurs de crue de barrage par exemple) des résultats, sinon plus exacte du moins plus cohérents, que ceux calculés à l'initiative de chacun.

M. SNEYERS craint que l'emploi d'un type donné de loi pour décrire les crues d'un groupe régional de rivières soit dangereux. Le comportement de chaque rivière est lié à celui de ses affluents et il est peu probable que l'influence totale des affluents soit la même pour chaque rivière.

En réalité, on est donc conduit, dans ce cas, à faire des ajustements en utilisant une loi "transformée" de la loi normale ou de la loi des valeurs extrêmes au moyen d'une transformation hyperbolique qui a l'avantage d'être, à la limite, une loi linéaire.

On peut démontrer théoriquement que, si plusieurs lois de probabilité concourent à déterminer une loi finale, cette loi finale est le produit des diverses lois de probabilité distinctes et, à l'aide de cette hypothèse, on peut voir qu'on aura un comportement asymptotiquement linéaire dans les petites valeurs et dans les grandes valeurs.

L'hétérogénéité de l'effet des affluents rend, de ce fait, inadéquate l'adoption d'une forme unique de loi pour un réseau hydrographique déterminé.

M. le Président croit que seule l'expérience dira si la réelle difficulté signalée par M. SNEYERS est dirimante en pratique.

M. Dino TONINI croit que du point de vue pratique, la proposition de M. le Président est bonne car l'ingénieur doit faire des calculs pour certaines régions et utiliser des formules ou des lois qui ne sont pas les mêmes et qui aboutissent à des résultats pratiques très divergents.

Les études de M. GUILLOT présentent un très grand intérêt au point de vue scientifique. Mais jusqu'à présent, on n'a jamais été dans les applications au-delà de la loi Galton-Gibrat.

On pourrait faire des comparaisons et arriver même à une loi très simple quelles que soient les régions. Ce serait une question très intéressante à suivre.

M. le Président observe qu'il restera à déterminer le type loi qui sera choisi comme norme dans telle ou telle région : cela représente un gros travail d'analyse d'une foule de données souvent disparates.

M. GUILLOT indique qu'il existe des papiers gauss-arithmétiques et gauss-logarithmiques parce que, il y a un siècle environ, on a pris la peine de tabuler d'une manière correcte la loi de Gauss et tout le monde les utilise parce qu'il est plus commode de prendre ce qui existe. Le jour où l'on aura fait des tables de lois d'Halphen qui seront commodes et faciles à manipuler, ou même des papiers Halphéno-logarithmiques, peut-être y aura-t-il des utilisateurs nombreux. Ce ne sera toujours là qu'un schéma numérique commode et non une loi physique.

M. le Président remercie à nouveau M. GUILLOT pour son remarquable exposé qui a donné à chacun l'occasion d'approfondir un problème encore loin d'être parfaitement résolu en pratique.