

Loi des fuites

Pierre RIBSTEIN
Hydrologue à l'ORSTOM

RÉSUMÉ

Cette étude de la Loi des fuites comprend une partie théorique avec, en particulier, les différentes méthodes d'estimation des deux paramètres de la loi et les problèmes de troncature. Les résultats théoriques sont illustrés par trois applications et par les sous-programmes Fortran.

ABSTRACT : « Law of leaks »

The interest of this law is the possibility to affect a not-null probability to the zero-value. It is a mixed-law, composed with exponential law and Poisson's law.

This study mainly explains how to estimate the statistical parameters and the problem of truncature. After the theory aspect, three applications relating to hydrology are given and the subprograms in Fortran are listed in annex.

Cette étude a été menée dans le cadre de l'ORSTOM, à la suite d'une étude préalable de M. Brunet-Moret et grâce à son aide constante.

A la base de cette étude, il y a un document que nous avons beaucoup utilisé : la thèse de C. Babusiaux, présentée à la faculté des Sciences de Paris en 1969 : « Etude statistique de la loi des fuites », thèse effectuée dans le cadre de la section Hydrologie de l'E.D.F. et suivie par M. Bernier.

SOMMAIRE

Introduction

1. *Présentation de la loi*
2. *Etude théorique*
 - 2.1. *Fonction de densité*
 - 2.1.1. *Expression mathématique*
 - 2.1.2. *Représentation de la fonction de densité*
 - 2.2. *Fonction de répartition*
 - 2.2.1. *Expression mathématique*
 - 2.2.2. *Inversion de la fonction de répartition*
 - 2.2.3. *Représentation de la fonction de répartition*
 - 2.3. *Moments, cumulants, valeurs centrales*
 - 2.3.1. *Moments non centrés de la distribution*
 - 2.3.2. *Cumulants de la distribution*
 - 2.3.3. *Asymétrie et aplatissement de la distribution*

- 2.4. Estimation des paramètres de la loi non tronquée
 - 2.4.1. Méthode des moments
 - 2.4.2. Méthode utilisant n_0 et la moyenne
 - 2.4.3. Méthode utilisant n_0 , \bar{x} et K_2
 - 2.4.4. Méthode du maximum de vraisemblance
 - 2.4.5. Efficacité
- 2.5. Troncature de la loi
3. *Programme Fortran*
 - 3.1. Simulation d'une loi des fuites
 - 3.1.1. Présentation
 - 3.1.2. Fonctionnement
 - 3.2. Fonction F FUIITE (Y, XLAM, S)
 - 3.2.1. Présentation
 - 3.2.2. Fonctionnement
 - 3.3. Fonction V FUIITE (P, XLAM, S)
 - 3.3.1. Présentation
 - 3.3.2. Fonctionnement
 - 3.4. Subroutine NTFUIT (XLAM, S, N, X, XM)
 - 3.4.1. Présentation
 - 3.4.2. Fonctionnement
 - 3.5. Subroutine OPTIM (XLAM, S, N, N1, X, XH)
 - 3.5.1. Présentation
 - 3.5.2. Fonctionnement
4. APPLICATIONS
 - 4.1. Valeurs simulées; ajustement d'une loi des fuites avec troncature
 - 4.2. Comparaison des différentes méthodes d'estimation des paramètres et troncature d'un échantillon de pluies décennales
 - 4.3. Ajustement d'une loi tronquée à différents pas de temps

Conclusion

Bibliographie

Annexe 1 : Fonctions de Bessel

Annexe 2 : Sous-programmes Fortran

INTRODUCTION

L'étude que nous nous proposons d'effectuer est relative à la loi des fuites et, plus particulièrement, à son application en hydrologie.

Ce sujet doit être explicité; pourquoi une loi statistique de plus, alors que les hydrologues sont déjà perdus devant la grande quantité de lois offertes à leur réflexion ? Il s'avère que cette loi présente comme principale propriété d'être une loi mixte : la variable aléatoire étant positive ou nulle, une probabilité non nulle est attachée à la valeur zéro, tandis que la densité de probabilité est continue pour les valeurs strictement positives de la variable.

L'utilisation des lois classiques : Galton, Pearson, etc. oblige à un tronquage de la loi, ce qui entraîne des calculs supplémentaires plus ou moins complexes. Pour la loi des fuites, l'existence d'une fréquence d'apparition pour les valeurs nulles élimine l'opération de tronquage en fréquence, mais non la troncature de l'échantillon, du fait des erreurs d'observations pour les valeurs faibles. Cette propriété est intéressante car, dans ces études statistiques, l'hydrologue est souvent amené à traiter des échantillons présentant un nombre important de valeurs nulles : pluies moyennes journalières, débits moyens journaliers, ou même mensuels, etc.

La deuxième particularité se remarque lors de l'étude des cumulants; la variable aléatoire, définie comme somme de m variables aléatoires indépendantes tirées d'une population mère distribuée suivant une loi des fuites, de paramètres λ , s et x_0 suit une loi des fuites de paramètres $m\lambda$, s et mx_0 . Ceci peut nous permettre d'étudier des événements chronologiques, à différents pas de temps, tout en conservant comme loi statistique, la loi des fuites.

Dans cette étude, nous insisterons particulièrement sur les déterminations des valeurs des paramètres, car ce point est important dans la pratique.

La méthodologie d'utilisation d'une loi statistique est décrite par Brunet-Moret; ayant un échantillon homogène de valeurs observées (supposées sans erreurs systématiques), nous choisissons une forme mathématique que nous supposons être, a priori, la forme mathématique de la distribution de la population-mère, d'où provient l'échantillon. D'après cet échantillon, nous déterminons la valeur des paramètres qui ajustent à l'échantillon la forme mathématique de la population-mère. Les valeurs des paramètres résument (pour la forme mathématique choisie, avec les conditions et les possibilités offertes par cette forme), l'information contenue dans l'échantillon, et un point très important est de perdre le moins possible de cette information.

1. PRÉSENTATION DE LA LOI

Cette loi a été présentée de différentes façons suivant les auteurs et l'utilisation qu'ils voulaient en faire. Elle a son origine dans la théorie du renouvellement, qui est l'étude d'événements chronologiques de longueur aléatoire, et séparés par des durées aléatoires.

Il est important de garder toujours en mémoire les hypothèses de base :

- stationnarité : les propriétés statistiques de la série chronologique ne changent pas avec le temps.
- indépendance des événements chronologiques.

Le nom de « loi des Fuites » provient de l'utilisation de cette loi dans l'étude des distributions des débits de fuite sur une conduite de gaz par M. Morlat. Nous utiliserons la définition employée dans cette étude.

Soit $y \geq 0$ variable aléatoire qui suit une loi exponentielle simple (loi gamma 1) de paramètres d'échelle s , et de position 0 , de densité de probabilité :

$$f(y) = \frac{1}{s} \cdot e^{-\frac{y}{s}}$$

Soit l un nombre entier aléatoire suivant une loi de Poisson, de paramètre de forme $\lambda > 0$; la probabilité de la valeur l est :

$$p(l) = \frac{e^{-\lambda} \cdot \lambda^l}{l!}$$

La variable aléatoire $x = y_1 + y_2 + \dots + y_l$ somme d'un nombre aléatoire l de variables aléatoires indépendantes y suit une loi des fuites de paramètres de forme λ , d'échelle s , de position 0 .

REMARQUES

Le mode de construction impose la valeur 0 comme valeur inférieure de l'intervalle de définition des variables x et y . On peut imaginer une borne inférieure de l'intervalle de définition de x qui soit un paramètre de position x_0 différent de zéro.

On peut démontrer qu'il est équivalent de dire qu'un nombre d'événements $n(t)$ suit une loi de Poisson sur une durée t , ou que la durée T entre deux événements suit une loi exponentielle sur cette même base de temps t .

$n(t)$: nombre d'événements sur la période $(0, t)$.

T : durée entre deux événements, suit une loi exponentielle de paramètre de forme μ_T , soit :

$$f(T) = \frac{1}{\mu_T} e^{-\frac{T}{\mu_T}}$$

Soit $t_r = T_1 + T_2 + \dots + T_r$ somme de r variables suivant une loi exponentielle. La loi conditionnelle de t_r est une loi gamma incomplète, de densité de probabilité :

$$f(t_r) = \frac{\left(\frac{t_r}{\mu_T}\right)^{r-1} \cdot e^{-\frac{t_r}{\mu_T}}}{\mu_T \cdot (r-1)!}$$

d'où

$$\text{Prob}(t_r \leq t) = \int_0^t f(t_r) \cdot dt_r = F_{t_r}(t)$$

donc

$$\text{Prob}(n(t) < r) = \text{Prob}(t_r > t) = 1 - F_{t_r}(t)$$

$$\text{Prob}(n(t) = r) = \text{Prob}(n(t) < r+1) - \text{Prob}(n(t) < r) = F_{t_r}(t) - F_{t_{r+1}}(t)$$

Soit

$$\text{Prob}(n(t) = r) = \int_0^t \frac{\left(\frac{x}{\mu_T}\right)^{r-1} \cdot e^{-\frac{x}{\mu_T}}}{\mu_T \cdot (r-1)!} dx - \int_0^t \frac{\left(\frac{x}{\mu_T}\right)^r \cdot e^{-\frac{x}{\mu_T}}}{\mu_T \cdot r!} dx$$

En intégrant par partie, on obtient :

$$\text{Prob}(n(t) = r) = \frac{e^{-\frac{t}{\mu_T}} \cdot \left(\frac{t}{\mu_T}\right)^r}{r!}$$

En partant d'une distribution exponentielle pour la durée entre deux événements, avec comme paramètre de forme μ_T , nous avons démontré que le nombre d'événements $n(t)$, sur une période t , suit une loi de Poisson de paramètre $\left(\frac{t}{\mu_T}\right)$.

Dans la suite de l'étude, nous utiliserons la distribution de Poisson pour le nombre d'événements $n(t)$.

La présentation de la loi, décrite précédemment, est celle que M. Morlat a utilisée pour l'étude des fuites de gaz le long d'un réseau.

En effet, si l'on suppose que :

— le nombre de points de fuite le long d'un tronçon de conduite de longueur donnée suit une loi de Poisson, de paramètre λ ,

— le débit de fuite en un point suit une loi exponentielle de paramètre s ,

alors, le débit total sur le tronçon, Z , suit une loi des fuites, de paramètre de forme λ et de paramètre d'échelle s .

H. A. Einstein s'est intéressé à la loi de probabilité de la distance parcourue par une particule de sable au fond d'une rivière, pendant un temps t déterminé. Il a fait l'hypothèse de la stationnarité du flot; la distance étudiée est la projection de la distance réelle sur la surface du niveau de l'eau. Enfin, le temps pendant lequel la particule est en mouvement est négligeable par rapport au temps d'arrêt.

A partir des considérations suivantes :

— la distance entre deux arrêts successifs suit une loi exponentielle,

— la probabilité du nombre d'arrêts sur l'intervalle $(0, t)$ ne dépend que de t et suit une loi de Poisson, on en déduit que la distance x , parcourue pendant un intervalle de temps t , suit une loi des fuites.

L'application de la loi des fuites à la pluviométrie se fait à partir des hypothèses suivantes :

— le nombre d'événements pluie à une station suit une loi de Poisson (il est équivalent de dire que la durée entre deux pluies suit une loi exponentielle),

— la hauteur d'eau tombée au cours d'une pluie suit une loi exponentielle.

A partir de ces 2 hypothèses et des hypothèses de base décrites précédemment, nous en déduisons que la hauteur d'eau tombée pendant un intervalle de temps t suit une loi des fuites.

Le choix de l'intervalle de temps t minimum dépend des caractéristiques des pluies de la région considérée :

— au Sahel, on pourra étudier les pluies sur une base journalière,

— en France, on prendra plutôt un intervalle de temps de plusieurs jours, car il n'y a pas indépendance des pluies tombées deux jours consécutifs.

Comme la loi de Poisson convient à l'ajustement des faibles probabilités (événements rares), la loi des fuites s'appliquera lorsque, sur le pas de temps choisi, le nombre d'événements pluie indépendants est faible.

2. ÉTUDE THÉORIQUE

2.1. FONCTION DE DENSITÉ

2.1.1. Expression mathématique

y variable aléatoire positive, suit une loi exponentielle simple, de paramètre d'échelle s , de position o . et de densité de probabilité :

$$f(y) = \frac{1}{s} e^{-\frac{y}{s}}$$

l nombre entier aléatoire suit une loi de Poisson, de paramètre de forme $\lambda > 0$, de probabilité :

$$p(l) = \text{Prob} (N = l) = e^{-\lambda} \cdot \frac{\lambda^l}{l!}$$

$x = y_1 + y_2 + \dots + y_l$, somme d'un nombre $l > 0$ fixé de variables aléatoires indépendantes y , suit une loi gamma incomplète, de densité de probabilité :

$$f(x, l) = \frac{1}{(l-1)!} \cdot \left(\frac{x}{s}\right)^{l-1} \cdot \frac{e^{-\frac{x}{s}}}{s}$$

En travaillant sur la variable réduite $u = \frac{x}{s}$.

$$f(u, l) = \frac{1}{(l-1)!} \cdot u^{l-1} \cdot e^{-u}$$

La loi marginale de u , pour l prenant toutes les valeurs de 1 à l'infini, a pour densité de probabilité :

$$f(u) = \sum_{l=1}^{\infty} \frac{e^{-\lambda} \cdot \lambda^l}{l!} \cdot \frac{1}{(l-1)!} \cdot u^{l-1} \cdot e^{-u}$$

$$\text{soit } f(u) = e^{-\lambda} \cdot \lambda \cdot e^{-u} \left[1 + \frac{\lambda u}{2} + \dots + \frac{(\lambda u)^l}{l!(l+1)!} + \dots \right]$$

La série entre crochets s'exprime à l'aide de la fonction de Bessel modifiée d'ordre 1 :

$$I_1(2\sqrt{\lambda u}) = \sum_{n=0}^{\infty} \frac{(\sqrt{\lambda u})^{2n+1}}{n!(n+1)!} = \sqrt{\lambda u} + \frac{(\sqrt{\lambda u})^3}{1!2!} + \dots + \frac{(\sqrt{\lambda u})^{2i+1}}{i!(i+1)!} + \dots$$

La fonction de densité de la loi des fuites s'écrit donc :

pour $u = \frac{x}{s}$

$$F_0 = e^{-\lambda}$$

$$f(u) = e^{-\lambda} \cdot \lambda \cdot e^{-u} \cdot \frac{I_1(2\sqrt{\lambda u})}{\sqrt{\lambda u}}$$

La distribution de u est une loi de probabilité mixte avec :

λ : paramètre de forme, positif,

s : paramètre d'échelle, positif.

Avec le changement de variable $z = 2\sqrt{\lambda u} = 2\sqrt{\frac{\lambda x}{s}}$, la fonction de densité peut aussi s'écrire :

$$f(z) = e^{-\lambda} \cdot \frac{2\lambda}{z} \cdot e^{-z^2/4\lambda} \cdot I_1(z).$$

— dérivée première de la fonction de densité :

$$f'(u) = e^{-\lambda} \cdot \lambda \cdot e^{-u} \left[\left(\frac{\lambda}{2} - 1 \right) + \dots + \frac{(\lambda u)^i}{i!(i+1)!} \left(\frac{\lambda}{i+2} - 1 \right) + \dots \right]$$

— dérivée seconde de la fonction de densité :

$$f''(u) = e^{-\lambda} \cdot \lambda \cdot e^{-u} \left[\left(1 - \lambda + \frac{\lambda^2}{6} \right) + \dots + \frac{(\lambda u)^i}{i!(i+1)!} \left(1 - \frac{2\lambda}{i+2} + \frac{\lambda^2}{(i+2)(i+3)} \right) + \dots \right]$$

2.1.2. Représentation de la fonction de densité

De l'annulation des dérivées premières et secondes de $f(u)$, on déduit la forme de la fonction de densité suivant les valeurs de λ .

— $f(u)$ est une loi mixte : il y a discontinuité au niveau de la valeur 0 (sauf lorsque $\lambda = 1$). En effet, l'ordonnée à l'origine est $\lambda e^{-\lambda}$ alors que la fonction de densité, pour la valeur 0, est : $f(0) = F_0 = e^{-\lambda}$.

— $\lambda < (3 - \sqrt{3})$ courbe représentative en « J », pas de point d'inflexion, mode observé en $u = 0$.

— $(3 - \sqrt{3}) \leq \lambda < 2$ courbe représentative en « J », un point d'inflexion, mode observé en $u = 0$.

— $2 \leq \lambda < (3 + \sqrt{3})$ le mode est réel, un seul point d'inflexion.

— $\lambda \geq (3 + \sqrt{3})$ le mode est réel, deux points d'inflexion.

La courbe prend l'allure en cloche, d'autant plus prononcée que λ est grand. Ce résultat se retrouve lorsqu'on étudie la fonction caractéristique de la loi des fuites qui, pour $\lambda \rightarrow \infty$, tend vers la fonction caractéristique de la loi de Gauss.

La figure 1 illustre ces résultats; nous avons tracé la fonction $f(u)$ pour différentes valeurs de λ , après avoir écrit le programme correspondant à l'expression de $f(u)$ donnée au paragraphe 2.1.1.

Loi des fuites

Fonction de densité pour différentes valeurs de λ

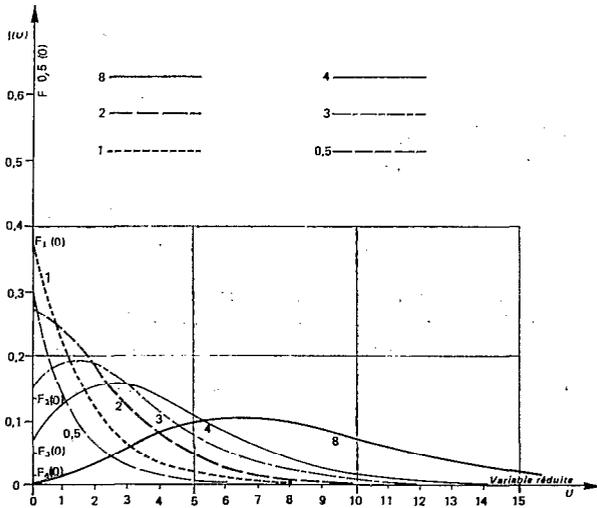


Fig. 1

Loi des fuites

Fonction de répartition pour différentes valeurs de λ

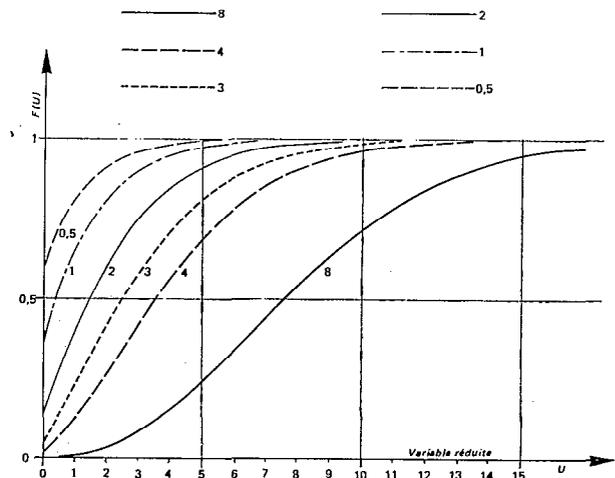


Fig. 2

2.2. FONCTION DE RÉPARTITION

2.2.1. Expression mathématique

$F(x)$ obtenu par intégration de la fonction densité en posant $u = \frac{x}{s}$ $F(x) = F_0 + \int_0^u f(u) du$

d'où
$$F(x) = e^{-\lambda} + e^{-\lambda} \int_0^u \lambda \cdot e^{-u} \left(\sum_{i=0}^{\infty} \frac{(\lambda u)^i}{i!(i+1)!} \right) du$$

soit
$$F(x) = e^{-\lambda} + e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \int_0^u \frac{e^{-u} \cdot u^i}{i!} du$$

On utilise l'équation :

$$\int_0^u \frac{e^{-u} \cdot u^i}{i!} du = \frac{-e^{-u} \cdot u^i}{i!} + \int_0^u \frac{e^{-u} \cdot u^{i-1}}{(i-1)!} du$$

soit
$$\int_0^u \frac{e^{-u} \cdot u^i}{i!} du = 1 - e^{-u} \sum_{j=0}^i \frac{u^j}{j!}$$

On en déduit :

$$F(x) = e^{-\lambda} + e^{-\lambda} \left[\sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \left(1 - e^{-u} \sum_{j=0}^i \frac{u^j}{j!} \right) \right]$$

Equation que l'on peut simplifier :

$$F(x) \text{ au non-dépassement} = 1 - e^{-\lambda} \cdot e^{-u} \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \sum_{j=0}^i \frac{u^j}{j!} \quad \text{avec } u = \frac{x}{s}$$

$$F_1(x) \text{ au dépassement} = e^{-\lambda} \cdot e^{-u} \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \sum_{j=0}^i \frac{u^j}{j!}$$

2.2.2. Inversion de la fonction de répartition

On ne peut pas calculer directement la valeur correspondant à une probabilité donnée P , les paramètres λ et s étant connus. La méthode qui nous a semblé la plus simple est de procéder de la façon générale; calculer \bar{u} , valeur moyenne, d'après le paramètre de forme connu, puis $F(\bar{u})$ probabilité correspondante. Si celle-ci est supérieure à la valeur donnée P de la probabilité à inverser, diviser \bar{u} par 2 : $u_1 = \frac{\bar{u}}{2}$. Calculer la probabilité correspondante et, si celle-ci est encore supérieure à P , diviser u_1 par 2 et ainsi de suite jusqu'à ce qu'une valeur u_i corresponde à une probabilité inférieure à P . On prend alors $u_j = \frac{u_i + u_{i-1}}{2} \dots$

Si la probabilité calculée pour \bar{u} est inférieure à la valeur donnée de la probabilité à inverser, multiplier \bar{u} par 2, etc. L'organigramme de cette méthode est présenté au paragraphe 3.3.

2.2.3. Représentation de la fonction de répartition

Dans le chapitre 3, nous décrivons le sous-programme relatif au calcul de la fonction de répartition. Ce sous-programme nous a permis de tracer la figure 2, pour des valeurs de λ déjà choisies dans la figure 1.

Nous voyons bien, sur la figure 2, l'intérêt de la loi des fuites : sans effectuer de tronquage en fréquence, on a une probabilité non nulle d'apparitions de valeurs nulles. Comme nous le verrons dans le chapitre « Estimation des paramètres », nous pouvons estimer le paramètre de forme λ à partir du nombre de valeurs nulles n_0 , si celui-ci est connu avec précision.

2.3. MOMENTS, CUMULANTS, VALEURS CENTRALES

2.3.1. Moments non centrés de la distribution

Le moment non centré d'ordre K de la variable réduite $u = \frac{x}{s}$ s'écrit :

$$m_{K,u} = \int_0^{\infty} u^K \cdot e^{-\lambda} \cdot \lambda \cdot e^{-u} \left[1 + \frac{\lambda u}{2} + \dots + \frac{(\lambda u)^i}{i!(i+1)!} + \dots \right] du$$

soit

$$m_{K,u} = e^{-\lambda} \cdot \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} \int_0^{\infty} \frac{u^{K+i-1} \cdot e^{-u}}{(i-1)!} du$$

Comme

$$\int_0^{\infty} \frac{e^{-u} \cdot u^{i+K}}{i!} du = \left[-\frac{e^{-u} \cdot u^{i+K}}{i!} \right]_0^{\infty} + \int_0^{\infty} \frac{e^{-u} \cdot (i+K) u^{i+K-1}}{i!} du = \frac{(i+K)!}{i!}$$

il vient :

$$m_{K,u} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} \cdot \frac{(i+K-1)!}{(i-1)!}$$

On en déduit le moment non centré d'ordre K de la variable non réduite :

$$m_{K,\lambda} = s^K \cdot m_{K,u} = s^K \cdot e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} \cdot \frac{(i+K-1)!}{(i-1)!}$$

La formule précédente nous permet de calculer :

- moyenne : $m_1 = \lambda s$
- moment non centré d'ordre 2 : $m_2 = \lambda s^2 (\lambda + 2)$
- moment non centré d'ordre 3 : $m_3 = \lambda s^3 (\lambda^2 + 6\lambda + 6)$
- moment non centré d'ordre 4 : $m_4 = \lambda s^4 (\lambda^3 + 12\lambda^2 + 36\lambda + 24)$

2.3.2. Cumulants de la distribution

La fonction caractéristique de première espèce de la loi des fuites se calcule par l'intégrale :

$$\Phi(t) = F_0 e^{it_0} + \int_0^{\infty} e^{itx} \cdot f(x) dx$$

C. Babusiaux, dans sa thèse, montre qu'on obtient après intégration terme à terme :

$$\Phi(t) = \exp \frac{its\lambda}{1 - its}$$

Soit $\Psi(t)$ la seconde fonction caractéristique :

$$\Psi(t) = \log \Phi(t) = \frac{it\lambda s}{1 - its} = it\lambda s + (it)^2 \cdot \lambda s^2 + \dots + (it)^n \cdot \lambda s^n + \dots$$

$\Psi(t)$ est la fonction génératrice des cumulants, d'où :

- moyenne : $K_1 = \lambda s$
- variance : $K_2 = 2\lambda s^2$
- moment centré d'ordre 3 : $K_3 = 6\lambda s^3$
- cumulant d'ordre 4 : $K_4 = 24\lambda s^4$
- cumulant d'ordre K : $K_K = K! \lambda s^K$

Le cumulant d'ordre $K \geq 1$ est K fois le cumulant de même ordre, d'une loi gamma incomplète de mêmes paramètres de forme, d'échelle et de position.

L'intérêt des cumulants provient de ce que, lorsque l'expression d'une variable est une combinaison linéaire de plusieurs variables indépendantes $X = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_i X_i$, chaque cumulant de X est la somme des cumulants de même ordre des variables indépendantes, pondérés comme suit :

$$K_r(X) = \alpha_1^r \cdot K_r(X_1) + \alpha_2^r \cdot K_r(X_2) + \dots + \alpha_i^r \cdot K_r(X_i)$$

Les valeurs des cumulants de la loi des fuites montrent que la variable aléatoire définie comme somme de m variables aléatoires indépendantes tirées d'une population mère distribuée suivant une loi des fuites de paramètres λ , s et x_0 , suit une loi des fuites de paramètres $m\lambda$, s et $m x_0$.

2.3.3. Asymétrie et aplatissement de la distribution

Le coefficient d'asymétrie est :

$$\gamma_1 = \frac{K_3}{K_3^{3/2}} = \frac{6\lambda s^3}{(2\lambda s^2) \sqrt{2\lambda s^2}} = \frac{3}{\sqrt{2\lambda}}$$

$\gamma_1 = \frac{3}{\sqrt{2\lambda}}$ ($\lambda > 0$) la distribution est asymétrique et plus étalée sur la droite que la loi normale; l'asymétrie diminue rapidement avec λ .

Le coefficient d'aplatissement est : $\gamma_2 = \frac{K_4}{K_2^2} = \frac{24\lambda s^4}{4\lambda^2 s^4} = \frac{6}{\lambda}$ ($\lambda > 0$) la distribution est d'autant moins aplatie par rapport à la loi normale que λ est plus grand.

Lorsque λ croît, les deux coefficients γ_1 et γ_2 tendent vers zéro et la loi des fuites tend vers une distribution de Gauss.

$$\text{Coefficient de variation : } CV = \frac{s\sqrt{2\lambda}}{s\lambda} = \sqrt{\frac{2}{\lambda}}$$

Remarque: Comparaison de l'asymétrie et de l'aplatissement d'une loi des fuites et d'une loi gamma incomplète de même coefficient de variation, et de même paramètre de position, nul.

	CV	γ_1	γ_2
Loi gamma incomplète	$\frac{1}{\sqrt{\gamma}}$	$\frac{2}{\sqrt{\gamma}}$	$\frac{6}{\gamma}$
Loi des fuites	$\sqrt{\frac{2}{\lambda}}$	$\frac{3}{\sqrt{2\lambda}}$	$\frac{6}{\lambda}$

De l'égalité des coefficients de variation, on déduit $\sqrt{\frac{2}{\lambda}} = \frac{1}{\sqrt{\gamma}}$ donc $\gamma = \frac{\lambda}{2}$.

Le coefficient d'asymétrie de la loi gamma s'écrit :

$$\gamma_1 = \frac{2\sqrt{2}}{\lambda} \left(> \frac{3}{\sqrt{2\lambda}} \right)$$

La loi des fuites est plus symétrique que la loi gamma incomplète, à même coefficient de variation.

Le coefficient d'aplatissement de la loi gamma s'écrit :

$$\gamma_2 = \frac{12}{\lambda} \left(> \frac{6}{\lambda} \right)$$

La loi des fuites est plus aplatie que la loi gamma incomplète à même coefficient de variation.

2.4. ESTIMATION DES PARAMÈTRES DE LA LOI NON TRONQUÉE

La valeur du paramètre de position est connue a priori et prise égale à 0, sans perte de généralité.

Le nombre n_0 d'observations $x = 0$ est connu sans équivoque, dans l'échantillon de taille n : c'est-à-dire que, même si l'observation x_i a une valeur tellement petite qu'on peut la prendre nulle pour les calculs, on sait qu'elle est différente de zéro et qu'on n'a pas compté cette observation dans le nombre n_0 d'observations $x = 0$.

Dans sa thèse, C. Babusiaux décrit plusieurs méthodes de calcul des paramètres, à partir des moments et du nombre de valeurs nulles n_0 . Après avoir défini la notion d'efficacité, nous comparerons chacune de ces méthodes à la méthode du maximum de vraisemblance qui, elle, a par définition une efficacité de 1.

Pour toutes ces méthodes, nous faisons les hypothèses de base suivantes, selon Brunet-Moret :

— Les valeurs observées correspondent à des variables aléatoires, indépendantes, connues sans erreur d'observation: l'échantillon observé est représentatif d'une population mère distribuée suivant la loi des fuites.

— Si les variables observées ne sont pas strictement indépendantes, on suppose que la taille de l'échantillon est suffisamment grande (suivant le degré de dépendance) pour qu'il reste représentatif de la population-mère.

Pour toutes les méthodes utilisant les moments calculés sur l'échantillon, on fait une autre hypothèse :

— Si les variables observées sont entachées d'erreurs d'observations, on suppose que ces erreurs sont aléatoires, indépendantes entre elles, indépendantes des valeurs numériques des observations, distribuées avec une moyenne nulle, un moment centré d'ordre 3 nul, une variance constante (homoscédasticité) et petite par rapport à la variance de l'échantillon.

2.4.1. Méthode des moments

$$\text{moyenne } \bar{x} = \frac{\sum_{i=0}^n x_i}{n} = \lambda s$$

$$\text{variance } K_2 = \frac{\sum x_i^2 - (\sum x_i) \cdot \bar{x}}{(n-1)} = 2\lambda s^2$$

On en déduit les paramètres

$$s = \frac{K_2}{2\bar{x}} \quad \lambda = \frac{2\bar{x}^2}{K_2}$$

2.4.2. Méthode utilisant n_0 et la moyenne

La proportion d'observations nulles sans équivoque est :

$$\frac{n_0}{n} = e^{-\lambda}$$

d'où

$$\lambda = \text{Log} \frac{n}{n_0} \quad s = \frac{\bar{x}}{\lambda}$$

2.4.3. Méthode utilisant n_0 , \bar{x} et K_2

L'efficacité des deux méthodes précédentes variant de manière opposée en fonction de λ , C. Babusiaux a cherché une combinaison linéaire des deux estimateurs, conservant une efficacité suffisante quelque soit λ .

En conclusion, les déterminations des paramètres sont :

$$s = \left(1 - (2 - \sqrt{2}) \sqrt{\frac{n_0}{n}} \right) \left(\frac{K_2}{2\bar{x}} - \frac{\bar{x}}{\text{Log} \frac{n}{n_0}} \right) + \frac{\bar{x}}{\text{Log} \frac{n}{n_0}}$$

$$\lambda = \left(1 - (2 - \sqrt{2}) \sqrt{\frac{n_0}{n}} \right) \left(\frac{2\bar{x}^2}{K_2} - \text{Log} \frac{n}{n_0} \right) + \text{Log} \frac{n}{n_0}$$

Les 3 méthodes, décrites ci-dessus, sont très faciles à mettre en œuvre. Une quatrième méthode, décrite par C. Babusiaux, utilisant $\sqrt{\bar{x}}$, permet une estimation des paramètres efficace pour toutes les valeurs de λ , mais demande des calculs plus complexes.

2.4.4. Méthode du maximum de vraisemblance

En utilisant la méthode du maximum de vraisemblance, on ajoute l'hypothèse de base suivante :
Si les variables observées sont entachées d'erreurs d'observation (erreurs non systématiques qui, elles, se répercuteront sur le paramètre de position et/ou le paramètre d'échelle), la méthode peut être utilisée :

— soit lorsque l'erreur à craindre sur les valeurs observées est répartie avec une moyenne nulle et une variance indépendante de la valeur qui aurait dû être observée ;

— soit lorsqu'elle est répartie avec une moyenne nulle et un écart-type proportionnel à la valeur qui aurait dû être observée.

Dans ces deux cas, on ne suppose pas que la fonction de répartition de l'erreur soit gaussienne, mais qu'elle conserve la même formulation mathématique avec les mêmes valeurs des paramètres de forme quelle que soit la valeur observée.

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

La vraisemblance est une quantité proportionnelle au produit des densités de probabilités.

$$\text{pour } u = \frac{x}{s} \quad \mathcal{L} = e^{-\lambda n_0} \cdot \prod_1^{n-n_0} \left(e^{-\lambda} \cdot \frac{\lambda}{s} \cdot e^{-u} \cdot \frac{I_1(2\sqrt{\lambda u})}{\sqrt{\lambda u}} \right)$$

(produit pour les $(n - n_0)$ valeurs $x_i \neq 0$)

soit :

$$\text{Log } \mathcal{L} = -\lambda n + \frac{(n - n_0)}{2} \text{Log } \frac{\lambda}{s} - \sum_1^{n-n_0} \frac{x_i}{s} - \frac{1}{2} \sum_1^{n-n_0} \text{Log } x_i + \sum_1^{n-n_0} \text{Log } I_1 \left(2\sqrt{\frac{\lambda x_i}{s}} \right)$$

Pour maximiser \mathcal{L} , on annule les dérivées partielles de $\text{Log } \mathcal{L}$ par rapport aux différents paramètres, ce qui donne le système d'équations :

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \lambda} = -n + \frac{n - n_0}{2\lambda} + \sum_1^{n-n_0} \frac{I_1'(z_i)}{I_1(z_i)} \cdot \frac{z_i}{2\lambda} = 0 \quad (1) \quad \frac{dz_i}{d\lambda} = \frac{z_i}{2\lambda}$$

$$\text{avec } z_i = 2\sqrt{\frac{\lambda x_i}{s}}$$

$$\frac{\partial \text{Log } \mathcal{L}}{\partial s} = -\frac{(n - n_0)}{2s} + \sum_1^{n-n_0} \frac{x_i}{s^2} - \sum_1^{n-n_0} \frac{I_1'(z_i)}{I_1(z_i)} \cdot \frac{z_i}{2s} = 0 \quad (2) \quad \frac{dz_i}{ds} = -\frac{z_i}{2s}$$

Par combinaison membre à membre des équations (1) et (2), on obtient :

$$\sum_1^{n-n_0} \frac{x_i}{s^2} - \frac{n\lambda}{s} = 0 \quad \text{soit } \boxed{\bar{x} = \lambda s.}$$

Nous voyons donc que la moyenne est un estimateur du maximum de vraisemblance.

Les propriétés des fonctions de Bessel nous permettent d'écrire :

$$I_1'(z) = I_0(z) - \frac{1}{z} \cdot I_1(z).$$

En utilisant cette propriété dans l'équation (1), il en résulte :

$$-n + \frac{1}{2} \sum_1^{n-n_0} \frac{z_i \cdot I_0(z_i)}{\lambda \cdot I_1(z_i)} = 0$$

En remarquant $\frac{z_i}{2\lambda} = \sqrt{\frac{x_i}{\bar{x}}}$, on obtient finalement :

$$\sum_1^{n-n_0} \frac{I_0(z_i)}{I_1(z_i)} \cdot \sqrt{x_i} = n\sqrt{\bar{x}}$$

Si λ tend vers 0, $\frac{I_0(z_i)}{I_1(z_i)}$ tend vers $+\infty$, quelque soit x_i ; si λ tend vers $+\infty$, $\frac{I_0(z_i)}{I_1(z_i)}$ tend vers 1, quelque soit x_i .

La décroissance de $\frac{I_0(z_i)}{I_1(z_i)}$ étant monotone, quelque soit x_i , lorsque λ croît de 0 à $+\infty$, et comme $\sum_1^{n-n_0} \sqrt{x_i}$ est nécessairement inférieur à $n\sqrt{\bar{x}}$, l'équation implicite contenant λ possède une solution et une seule, positive. La solution sera déterminée par itération avec $s = \frac{\bar{x}}{\lambda}$ recalculé à chaque nouvelle valeur de λ .

2.4.5. Efficacité

C. Babusiaux cite la définition classique exposée dans l'ouvrage de Wilks, de l'efficacité d'un estimateur :

Soit
$$V = \begin{bmatrix} V(\lambda) & \text{cov}(\lambda, s) \\ \text{cov}(\lambda, s) & V(s) \end{bmatrix}$$

matrice des variances et covariance de λ et s .

Soit
$$I = \begin{bmatrix} V\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \lambda}\right) & \text{cov}\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \lambda}, \frac{\partial \text{Log } \mathcal{L}}{\partial s}\right) \\ \text{cov}\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \lambda}, \frac{\partial \text{Log } \mathcal{L}}{\partial s}\right) & V\left(\frac{\partial \text{Log } \mathcal{L}}{\partial s}\right) \end{bmatrix}$$

matrice des variances et covariance de $\frac{\partial \text{Log } \mathcal{L}}{\partial \lambda}$ et $\frac{\partial \text{Log } \mathcal{L}}{\partial s}$

I est aussi appelée matrice d'information.

L'efficacité d'un estimateur sera définie par la courbe $E = \frac{1}{|I| \cdot |V|}$.

D'après cette définition, on voit que l'efficacité des estimateurs du maximum de vraisemblance est égale à 1. Pour les autres méthodes, le but recherché est de déterminer les estimateurs pour lesquels la courbe E est la plus proche de $E = 1$.

Les résultats qui suivent sont issus de la thèse de C. Babusiaux.

— déterminant de la matrice d'information :

$$|I| = \frac{2n^2}{s^2} \left[-\lambda + \frac{1}{2} - \frac{1}{4\lambda} + \frac{e^{-\lambda}}{4\lambda} + \frac{e^{-\lambda}}{4} \cdot J(\lambda) \right]$$

avec
$$J(\lambda) = \int_0^\infty e^{-u} \cdot \frac{\left[1 + 3 \cdot \frac{\lambda u}{1! 2!} + \dots + \frac{2i+1}{i!(i+1)!} \cdot (\lambda u)^i + \dots \right]^2}{\left[1 + \frac{\lambda u}{1! 2!} + \dots + \frac{(\lambda u)^i}{i!(i+1)!} + \dots \right]} du$$

$J(\lambda)$ a été tabulée par l'auteur citée ci-dessus.

— efficacité de la méthode des moments : $|V_1| = \frac{s^2}{n^2} (3 + 4\lambda)$

d'où
$$E_1 = \frac{1}{|I| \cdot |V_1|} = \frac{1}{2 \cdot (3 + 4\lambda) \left(-\lambda + \frac{1}{2} - \frac{1}{4\lambda} + \frac{e^{-\lambda}}{4\lambda} + \frac{e^{-\lambda}}{4} \cdot J(\lambda) \right)}$$

Cette équation est représentée par la courbe 1 de la figure 3. L'efficacité tend vers 1 lorsque λ tend vers l'infini, mais on remarque aussi que E décroît rapidement pour $\lambda < 5$, c'est-à-dire que lorsque les valeurs observées sont faibles, la méthode des moments ne peut plus s'appliquer, en particulier, si il y a beaucoup de valeurs nulles.

— efficacité de la méthode utilisant \bar{x} et $\frac{n_0}{n}$

$$|V_2| = \frac{\lambda^2}{n^2} \left(\frac{2}{\lambda} (e^\lambda - 1) - 1 \right) \text{ soit } E_2 = \frac{1}{2 \cdot \left(\frac{2}{\lambda} (e^\lambda - 1) - 1 \right) \left(-\lambda + \frac{1}{2} - \frac{1}{4\lambda} + \frac{e^{-\lambda}}{4\lambda} + \frac{e^{-\lambda}}{4} J(\lambda) \right)}$$

Cette équation représentée par la courbe 2 de la figure 3 montre que cette méthode est efficace lorsque $\lambda < 3$, c'est-à-dire lorsque le pourcentage de valeurs nulles est assez important.

D'autre part, en hydrologie, la connaissance des faibles valeurs est difficile et ne nous permet pas d'avoir une bonne précision sur le rapport $\frac{n_0}{n}$. Cette méthode ne s'appliquera donc que lorsque le nombre de valeurs nulles est assez important et connu avec précision.

— efficacité de la méthode utilisant \bar{x} , K_2 et $\frac{n_0}{n}$

$$|V_3| = -1 + 4p + 4p^2\lambda + (e^\lambda - 1) \frac{p^2 + 1 - 2p}{\lambda} \text{ avec } p = 1 - (2 - \sqrt{2}) e^{-\lambda/2}$$

d'où :

$$E_3 = \frac{1}{|I| \cdot |V_3|} = \frac{1}{2 \cdot \left(-1 + 4p + 4p^2\lambda + (e^\lambda - 1) \left(\frac{p^2 + 1 - 2p}{\lambda} \right) \right) \left(-\lambda + \frac{1}{2} - \frac{1}{4\lambda} + \frac{e^{-\lambda}}{4\lambda} + \frac{e^{-\lambda}}{4} J(\lambda) \right)}$$

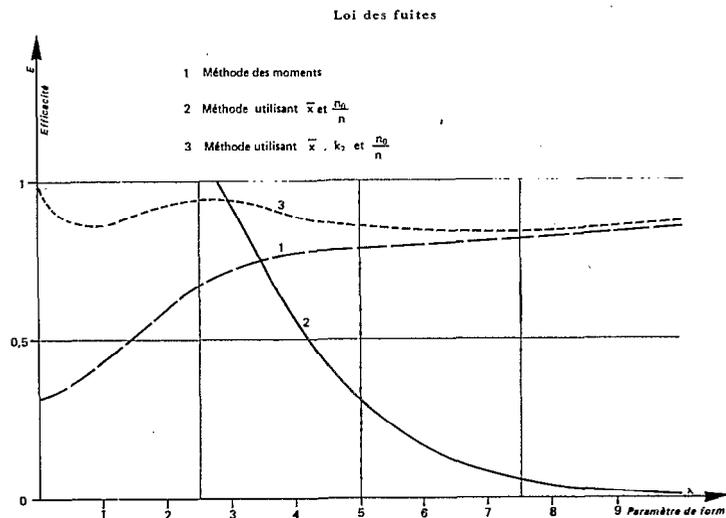


Fig. 3

Comme dans le cas précédent, le problème de la bonne connaissance de $\frac{n_0}{n}$ se pose. Mais on peut remarquer sur la courbe 3 de la figure 3 que cette méthode est efficace quelle que soit la valeur de λ , et elle ne demande pas de gros moyens de calcul comme la méthode du maximum de vraisemblance.

En conclusion, dans le cas où il n'y a pas à faire de troncature de l'échantillon, et en l'absence d'ordinateur, c'est la méthode mixte, utilisant \bar{x} , K_2 et $\frac{n_0}{n}$ qui nous semble la plus appropriée, si toutefois le nombre de valeurs nulles n_0 est connu assez précisément.

2.5. TRONCATURE DE LA LCI

On est obligé de considérer des troncatures des observations lorsqu'on ne peut plus distinguer entre les valeurs $x = x_0$ et les valeurs $x = x_0 + \varepsilon$. Brunet-Moret cite le cas, entre autres, des observations pluviométriques journalières où ne sont notées que les hauteurs supérieures à 0,1 mm ; nous connaissons le nombre des observations supérieures ou égales à $x_0 + \varepsilon$ ainsi que les valeurs correspondantes de la variate, mais non le nombre d'observations qui auraient dû être faites entre x_0 et $x_0 + \varepsilon$, ni leur valeur numérique ; $x_h = x_0 + \varepsilon$ est le seuil de troncature.

Dans ce cas, les méthodes d'estimation des paramètres à partir des moments ne peuvent plus s'appliquer. On utilisera la méthode du maximum de vraisemblance en définissant la vraisemblance par :

$$\mathcal{L} = \left[F(x_h) \right]^{n_1} \prod_{i=1}^n f(x_i)$$

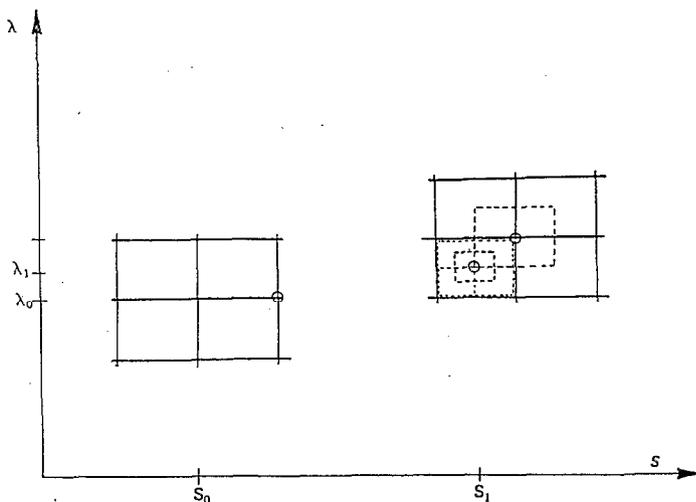
avec : x_h : seuil de troncature

n : nombre d'observations $x_i \geq x_h$

n_1 : nombre d'observations $x_i < x_h$

En utilisant la définition de la fonction de densité, on obtient :

$$\text{Log } \mathcal{L} = n_1 \text{Log } F(x_h) - n\lambda + n \text{Log } \frac{\lambda}{s} - \sum_{i=1}^n \frac{x_i}{s} + \sum_{i=1}^n \text{Log} \left[\frac{I_1(2\sqrt{\lambda u})}{\sqrt{\lambda u}} \right]$$



Recherche du maximum de vraisemblance

Fig. 4

« Classiquement », on trouve le maximum de la vraisemblance en annulant les dérivées partielles de $\text{Log } \mathcal{L}$ par rapport aux différents paramètres. Cette méthode nous a conduit à un système d'équations non linéarisables, trop compliqué à résoudre.

Nous avons préféré rechercher directement le maximum de $\text{Log } \mathcal{L}$ par optimisation. Le sous-programme proposé découpe le plan (λ, s) en 4 mailles, soit 9 nœuds, centrées sur les valeurs de λ_0 et s_0 fournies par la méthode des moments. Il calcule ensuite $\text{Log } \mathcal{L}$ en chaque nœud et recentre le maillage sur la valeur maximum trouvée. Si la valeur maximum se trouve au centre, il divise par 2 la taille des mailles et recommence la recherche du maximum jusqu'à obtenir une taille de maille suffisamment petite, c'est-à-dire une bonne précision sur l'estimation des paramètres λ_1 et s_1 . Le principe de la méthode est illustré par la figure 4.

3. PROGRAMMES FORTRAN

3.1. SIMULATION D'UNE LOI DES FUITES

Ce programme a été utilisé pour mettre au point les programmes d'estimation des paramètres de la loi des fuites. Nous avons simulé un nombre donné de valeurs tirées d'une population-mère, suivant une loi des fuites de paramètres λ et s choisis a priori, et nous avons ensuite utilisé les sous-programmes d'estimation pour retrouver ces valeurs.

3.1.1. Présentation

Ce programme permet de simuler un échantillon de N valeurs supérieures ou égales au seuil de troncature XH , et tirées d'une population-mère distribuée suivant une loi des fuites de paramètres de forme XLAM et d'échelle S .

Tel qu'il est présenté en annexe, le programme calcule les valeurs XLAM et S sur l'échantillon avec troncature. Il faut, pour cela, lui adjoindre les sous-programmes OPTIM et VRAIS d'optimisation du maximum de vraisemblance et le sous-programme FFUITE qui calcule la probabilité au non-dépassement pour une valeur Y donnée de la variate.

Cartes de données :

- 1^{re} carte : nombre au hasard IX initial ; il est recommandé de le prendre impair et de 7 à 8 chiffres.
- 2^e carte : N taille de l'échantillon ; vu les dimensions demandées, le programme ne permet pas de simuler plus de 400 valeurs.
- 3^e carte : lecture du paramètre de forme XLAM, du paramètre d'échelle S et du seuil de troncature XH choisis ; XH sera choisi à partir de la moyenne calculée par $XM = XLAM * S$.

3.1.2. Fonctionnement

Les valeurs ont été simulées en utilisant la méthode de Monte-Carlo, c'est-à-dire par anamorphose entre une distribution continue uniforme sur le segment $(0, 1)$ et la fonction de répartition choisie. La génération de nombres au hasard est faite suivant le schéma proposé par IBM, par la méthode de congruence multiplicative suivante :

$$X_i = 65539 * X_{i-1} \text{ (modulo } 2147483647)$$

Pour le calcul de chaque valeur, le programme peut être décomposé en deux parties :

- tirage du nombre d'événements L sur la durée t dans une loi de Poisson de paramètre XLAM. On construit la loi de Poisson par la formule $FL(l) = e^{-\lambda} + \sum_{i=1}^{l-1} \frac{e^{-\lambda} \cdot \lambda^i}{i!}$. On tire ensuite un nombre au hasard F compris entre 0 et 1. On cherche la valeur L qui a une probabilité F d'apparition, ce qui s'écrit par l'équation :

$$FL(L-1) \leq F < FL(L)$$

on va ainsi simuler : $L = L - 1$ événements ... (il y a un décalage de 1 car, pour 0 événements, on a $FL(1) = e^{-\lambda}$).

— tirage de L événements dans une loi exponentielle de paramètre s . la densité de probabilité d'une loi exponentielle est $F = \frac{1}{s} e^{-\frac{y_i}{s}}$. On en déduit : $y_i = s \cdot \text{LOG}(1/F.s)$.

Ayant tiré F au hasard compris entre 0 et 1, on en déduit la valeur y_i par l'équation précédente.

On en déduit finalement la valeur X , tirée d'une population-mère suivant une loi des fuites de paramètres λ et s , par l'équation :

$$X = \sum_{i=1}^L y_i.$$

3.2. FONCTION FFUITE (Y, XLAM, S)

3.2.1. Présentation

Ce sous-programme calcule la probabilité au non-dépassement FFUITE correspondant à la valeur Y de la variate. Arguments à fournir : valeur $Y \geq 0$ de la variable, paramètre de forme XLAM positif et paramètre d'échelle S positif.

3.2.2. Fonctionnement

Ce sous-programme utilise la formule présentée au paragraphe 2.2. Pour une valeur quelconque de $u = \frac{x}{s}$, la valeur de la fonction de répartition est le complément à 1 du produit ($e^{-\lambda} \cdot e^{-u}$) par une série dont chaque terme T est le produit d'une expression V_i par une série limitée $W(i)$.

Ces séries limitées ne diffèrent que par le rang du dernier terme à calculer : leur premier terme W_0 étant 1, un terme peut se déduire du précédent : $W_{j+1} = W_j \cdot \frac{u}{j}$. On arrête la sommation après avoir calculé le terme W_i de même rang i que V .

La première valeur V_0 de l'expression V_i multiplicatrice de la série limitée W est λ . Chaque valeur de cette expression se déduit de la précédente par $V_{i+1} = V_i \cdot \frac{\lambda}{i+1}$.

On peut arrêter la sommation de la série T lorsque le terme $V_i \cdot W(i)$ devient suffisamment petit par rapport

à la somme $\sum_{j=0}^i V_j \cdot W(j)$ des $(i+1)$ termes calculés, par exemple $\frac{V_i \cdot W(i)}{\sum_{j=0}^i V_j \cdot W(j)} < 10^{-6}$.

3.3. FONCTION VFUITE (P, XLAM, S)

3.3.1. Présentation

Ce sous-programme calcule la valeur VFUITE de la variable correspondant à la probabilité P au non-dépassement. Il faut lui adjoindre le sous-programme FFUITE, de calcul de la probabilité au non-dépassement.

Arguments à fournir : Probabilité P au dépassement ($0 < P < 1$), paramètre de forme XLAM positif, paramètre d'échelle S positif.

3.3.2. Fonctionnement

Nous avons décrit le principe du calcul dans le paragraphe 2.2.2. Nous pouvons l'illustrer par l'organigramme de la figure 5. On voit que le programme s'arrête, soit lorsque la probabilité au non-dépassement, F , de la variable

cherchée VFUITE est suffisamment proche de P, par exemple $\frac{(F - P)}{P} < 10^{-6}$, soit lorsque le nombre de calculs est égal ou supérieur à 100. Dans ce dernier cas, le programme écrit, au-dessus de la valeur estimée, « convergence non atteinte ».

3.4. SUBROUTINE NTFUIT (XLAM, S, N, X, XM)

3.4.1. Présentation

Ce sous-programme détermine les valeurs des paramètres de forme XLAM et d'échelle S par la méthode du maximum de vraisemblance pour un échantillon sans troncature. Ce sous-programme est extrait d'un programme mis au point par Brunet-Moret qui, sur un échantillon donné, teste neuf lois statistiques dont la loi des fuites. Il n'y a aucun sous-programme à adjoindre.

Arguments à fournir : la taille N de l'échantillon, les valeurs X de l'échantillon en vecteur X(N) et la valeur de la moyenne qui doit être calculée dans le programme principal.

3.4.2. Fonctionnement

Ce sous-programme utilise les équations du maximum de vraisemblance décrites au paragraphe 2.4.4. Le terme $\frac{I_0(z_1)}{I_1(z_1)} \cdot \sqrt{x_1}$ est déterminé à partir de l'approximation des fonctions de Bessel qui est donnée dans l'annexe 1.

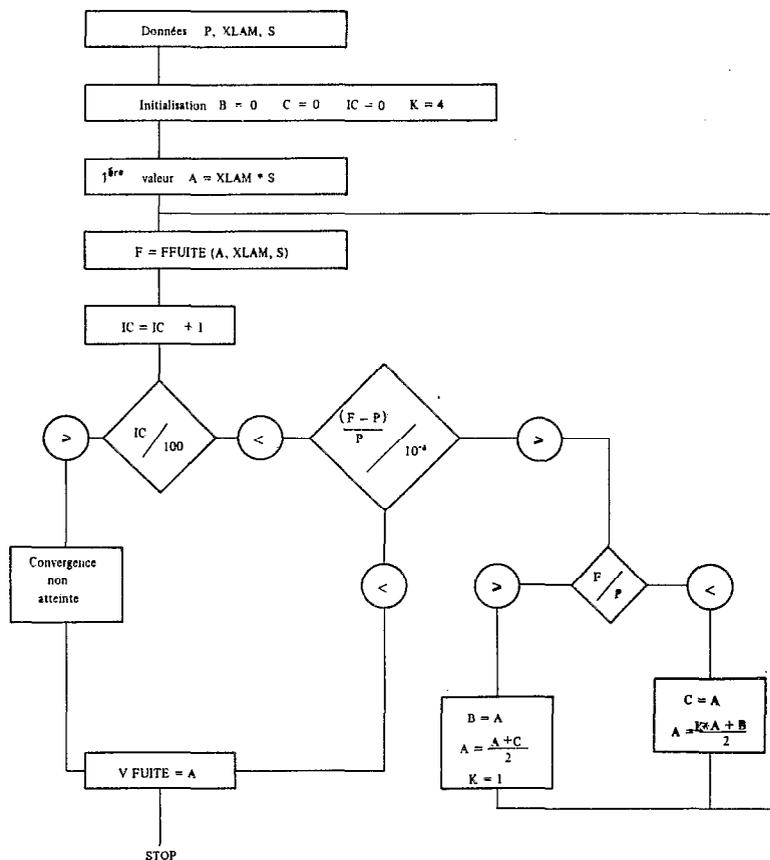


Fig. 5

3.5. SUBROUTINE OPTIM (XLAM, S, N, N1, X, XH)

3.5.1. Présentation

Ce sous-programme permet de déterminer les valeurs des paramètres de forme XLAM et d'échelle S sur un échantillon avec troncature. Il peut être divisé en deux sous-programmes, comme c'est le cas dans le listing présenté en annexe. Subroutine OPTIM est alors le programme de discrétisation du plan (XLAM, S) correspondant à la figure 4 du paragraphe 2.5. Subroutine VRAIS calcule le logarithme de la fonction de vraisemblance pour chaque couple de valeurs XLAM et S donné par le programme d'optimisation.

Arguments à fournir : le seuil de troncature XH, le nombre de valeurs N1 inférieures à XH, le nombre de valeurs N, supérieures ou égales au seuil XH, les valeurs X de l'échantillon ($X \geq XH$) présentées en vecteur X(N).

Le sous-programme OPTIM nécessite aussi une première estimation des valeurs des paramètres de forme et d'échelle, qui doivent être calculées dans le programme principal par la méthode des moments par exemple.

En plus des sous-programmes OPTIM et VRAIS, il faut ajouter fonction FFUITE qui détermine la probabilité au non-dépassement, pour une valeur donnée.

3.5.2. Fonctionnement

La description de la méthode a été faite au paragraphe 2.5. Le programme présenté part d'une taille de maille de $\left(\frac{\lambda}{2,5}, \frac{s}{2,5}\right)$. A chaque fois que le maximum de vraisemblance se trouve au centre du maillage, la taille des mailles est divisée par 2.

Les valeurs initiales de 2,5 pour λ et s peuvent être changées sans grande influence sur le nombre de calculs effectués.

L'optimisation s'arrête, soit lorsque la précision sur les paramètres est assez grande, par exemple $\pm 1 \cdot E - 08$, soit lorsque le nombre de maillages calculés a dépassé 200.

4. APPLICATIONS

4.1. VALEURS SIMULÉES ; AJUSTEMENT D'UNE LOI DES FUITES AVEC TRONCATURE

Comme nous l'avons souligné au paragraphe 3.1., le programme de simulation a été mis au point pour permettre de tester les méthodes d'évaluation des paramètres. A partir d'une même valeur de départ, nous avons simulé 100, 200, puis 300 valeurs issues d'une population-mère suivant une loi des fuites, de paramètres de forme $\lambda = 1$ et d'échelle $s = 1$. Ces valeurs sont supérieures au seuil de troncature choisi $XH = 1,5$.

En utilisant ensuite le programme de calcul du maximum de vraisemblance pour une loi tronquée, nous avons recalculé ces valeurs λ et s sur les échantillons simulés. Nous avons obtenu les résultats suivants :

N = 100 valeurs $X_i \geq 1,5$			
nombre de valeurs < 1,5 : N1 = 253	$\lambda = 1,353$	$s = 0,814$	
N = 200 valeurs $X_i \geq 1,5$			
nombre de valeurs < 1,5 : N1 = 594	$\lambda = 1,082$	$s = 0,912$	
N = 300 valeurs $X_i \geq 1,5$			
nombre de valeurs < 1,5 N1 = 885	$\lambda = 1,053$	$s = 0,945$	

Le résultat est trivial; encore fallait-il l'obtenir : en augmentant le nombre de valeurs simulées, nous avons une nette progression des valeurs calculées vers les valeurs choisies $\lambda = 1$ et $s = 1$.

Il faut souligner enfin que nous avons obtenu des valeurs λ et s très variables suivant la valeur IX choisie comme départ de la simulation. Pour $N = 100$, λ obtenu a varié de 0,87 à 1,35.

4.2. COMPARAISON DES DIFFÉRENTES MÉTHODES D'ESTIMATION DES PARAMÈTRES ET TRONCATURE D'UN ÉCHANTILLON DE PLUIES DÉCADAIRES

Nous avons repris les données utilisées par C. Babusiaux dans sa thèse. Il s'agit de relevés de précipitations en mm d'eau sur une période de 10 jours, effectués à Chateaufort de Randon (Cantal).

Rangement des 86 observations					
Rang R	Hauteur mm	F (R-0,5)/N	Rang R	Hauteur mm	F (R-0,5)/N
15	< 0,1	0,1686	58	8,0	0,6686
16	0,1	0,1802	59	8,3	0,6802
17	0,2	0,1919	60	8,4	0,6919
19	0,3	0,2151	61	8,6	0,7035
21	0,4	0,2384	62	8,8	0,7151
23	0,5	0,2616	63	8,9	0,7267
25	0,9	0,2849	64	9,3	0,7384
26	1,0	0,2965	65	10,0	0,7500
27	1,1	0,3081	66	11,1	0,7616
33	1,7	0,3779	67	11,4	0,7733
34	2,4	0,3895	68	11,5	0,7849
37	2,7	0,4244	69	12,6	0,7965
39	2,9	0,4477	70	14,3	0,8081
40	3,0	0,4593	72	15,2	0,8314
41	3,3	0,4709	73	15,9	0,8430
42	3,4	0,4826	74	16,3	0,8547
43	3,5	0,4942	75	17,4	0,8663
44	3,7	0,5058	76	17,9	0,8779
45	4,0	0,5174	77	19,0	0,8895
46	4,1	0,5291	78	20,3	0,9012
47	4,3	0,5407	79	20,7	0,9128
48	4,4	0,5523	80	22,6	0,9244
49	4,7	0,5640	81	24,8	0,9360
51	5,0	0,5872	82	29,6	0,9477
52	5,7	0,5988	83	31,2	0,9593
53	6,1	0,6105	84	35,5	0,9709
54	6,5	0,6221	85	36,7	0,9826
55	6,6	0,6337	86	41,7	0,9942
56	7,1	0,6453			
moyenne : $\bar{x} = 7,35$ mm			variance: 86,8		

Sur cet échantillon, nous avons calculé les paramètres λ et s avec toutes les méthodes décrites précédemment. Pour comparer les ajustements obtenus, nous avons appliqué le test du χ_2 sur l'échantillon découpé en 8 classes de probabilité, comme suit :

Numéro de classe :	1	2	3	4	5	6	7	8
Limite supérieure de probabilité :	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Effectif théorique :	25,8	8,6	8,6	8,6	8,6	8,6	8,6	8,6

Estimation des paramètres λ et s , test du χ_2

Méthodes	λ	s	χ_2
Méthode des moments	1,247	5,899	12,29
Méthode utilisant $\frac{n_0}{n}$ et \bar{x}	1,609	4,571	5,16
Méthode utilisant $\frac{n_0}{n}$, \bar{x} et K_3	1,342	5,551	5,55
Méthode du maximum de vraisemblance	1,610	4,568	5,16

Pour $\nu = 8 - 2 - 1 = 5$ degrés de liberté, $\chi_2 = 5,16$ a une probabilité de 0,65 et $\chi_2 = 12,29$ une probabilité de 0,10 en théorie. Si on choisit comme seuil du risque d'adéquation $\alpha = 0,10$, cela signifie qu'on considère comme satisfaisant tout ajustement pour lequel $P(\chi_2) > 0,10$. Dans ce cas, on rejette l'ajustement obtenu par la méthode des moments, les trois autres ajustements étant satisfaisants.

Les résultats obtenus par le test du χ_2 se retrouvent dans l'étude de l'efficacité des différentes méthodes. Sur la figure 3 du paragraphe 2.4.5., on voit que l'efficacité de la méthode utilisant $\frac{n_0}{n}$ et \bar{x} est égale à 1, l'efficacité de la méthode mixte est proche de 1 et l'efficacité de la méthode des moments est faible, pour des valeurs de λ entre 1,2 et 1,6. L'efficacité de la méthode du maximum de vraisemblance est égale à 1, quel que soit λ .

Pour avoir une bonne représentation des valeurs significativement différentes de 0, nous avons effectué une troncature de l'échantillon avec comme seuil $XH = 1,01$ mm et 3,01 mm. Sur l'échantillon avec troncature, nous avons calculé les paramètres de forme λ et d'échelle s par la méthode d'optimisation du maximum de vraisemblance.

$$XH = 1,01 \text{ mm} \quad \lambda = 1,389 \quad s = 5,274$$

$$XH = 3,01 \text{ mm} \quad \lambda = 1,184 \quad s = 6,060$$

Dans le tableau suivant, nous présentons quelques points des distributions ajustées par le maximum de vraisemblance sans troncature, et par le maximum de vraisemblance avec les troncatures choisies précédemment.

Récurrence	Fréquence	Hauteur (mm)		
		Echantillon non tronqué	Troncature $XH = 1,01$ mm	Troncature $XH = 3,01$ mm
2	0,5	4,9	4,4	3,7
5	0,8	13,0	13,2	13,3
10	0,9	18,6	19,3	19,9
100	0,99	35,4	38,0	40,5
1000	0,999	50,8	55,3	59,8

En augmentant le seuil de troncature, on donne plus de poids aux fortes valeurs de l'échantillon et donc on diminue la valeur du paramètre de forme λ . En théorie, on a vu qu'une diminution de λ entraîne une augmentation du coefficient d'aplatissement γ , c'est-à-dire une augmentation de la « courbure » de la fonction de distribution.

Ces remarques expliquent les résultats précédents : une augmentation du seuil XH entraîne une diminution des valeurs autour de la médiane et une augmentation des valeurs extrêmes.

La figure 6 montre l'ajustement aux valeurs observées de la fonction de distribution calculée pour une troncature de 3,01 mm. En pointillés, nous avons représenté la fonction de distribution pour les valeurs inférieures au seuil de troncature, non prises en compte dans l'ajustement.

4.3. AJUSTEMENT D'UNE LOI TRONQUÉE A DIFFÉRENTS PAS DE TEMPS

Nous avons étudié les précipitations maximales indépendantes sur 1, 2, 3 et 4 jours consécutifs à la station « Le Raizet » de la Guadeloupe.

Données : 28 années de relevés, soit 10 227 jours = N

moyenne annuelle des précipitations = 1 793 mm.

Après critique des données de pluviométrie journalière, nous avons utilisé un programme de l'ORSTOM — POH 141 — qui range en classes de 5 mm les précipitations indépendantes supérieures à un seuil donné pour le nombre de jours consécutifs choisis.

Nous avons ensuite intégré la loi des fuites à un programme déjà existant — POH 142 — qui ajuste, sur les données fournies par le programme précédent, la loi de Galton (loi Gauss-logarithmique) et la loi de Goodrich (loi exponentielle généralisée), ces deux lois étant tronquées en fréquence.

Pour 1, 2, 3 et 4 jours consécutifs, nous avons cherché les seuils de troncature qui nous permettaient de sélectionner entre 100 et 300 valeurs. Il faut noter que le terme « seuil de troncature » n'a plus la même signification que précédemment ; on ne cherche pas le seuil pour lequel x_0 est différent de $x_0 + \varepsilon$, mais on veut obtenir un bon ajustement pour les valeurs fortes, ce qui entraîne un seuil élevé, pour éviter de donner trop de poids aux faibles valeurs et aussi, bien sûr, pour ne pas tester trop de données.

Pour les précipitations sur 1 et 2 jours, il s'est avéré que, pour la loi des fuites, les seuils les plus élevés donnaient un bon ajustement. A l'inverse, pour des pas de temps supérieurs, 3 et 4 jours, les seuils les plus bas donnaient les meilleurs résultats. Il semble donc que les populations de précipitation sur 1 et 2 jours n'ont pas les mêmes caractéristiques statistiques que les précipitations sur trois jours et quatre jours.

Cette remarque pourrait s'expliquer par la non-indépendance des précipitations en Guadeloupe sur des pas de temps trop faibles, inférieurs à deux jours.

Nous présentons, dans le tableau ci-dessous, les résultats pour les seuils qui nous donnent le meilleur ajustement de la loi des fuites.

Ajustement de la loi des fuites				
Pas de temps	Seuil	Nombre de valeurs retenues	Paramètre de forme λ	Paramètre d'échelle s
1	50	125	0,0954	23,70
2	60	126	0,1737	29,33
3	50	234	0,3718	27,6
4	50	262	0,4991	29,26

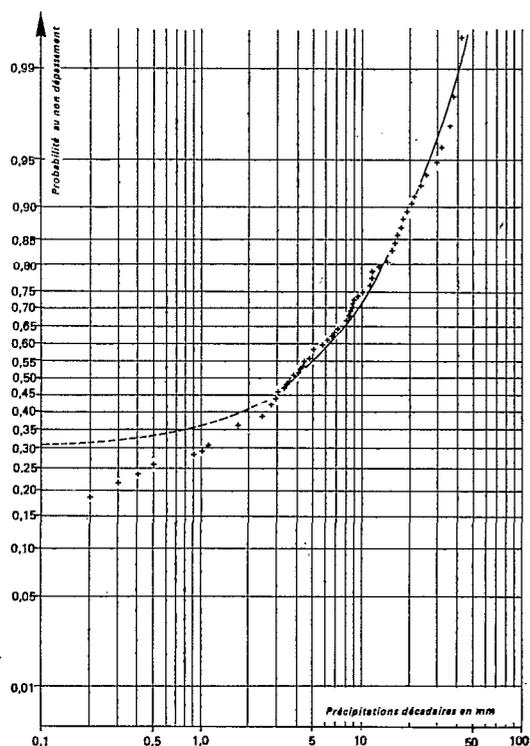


Fig. 6

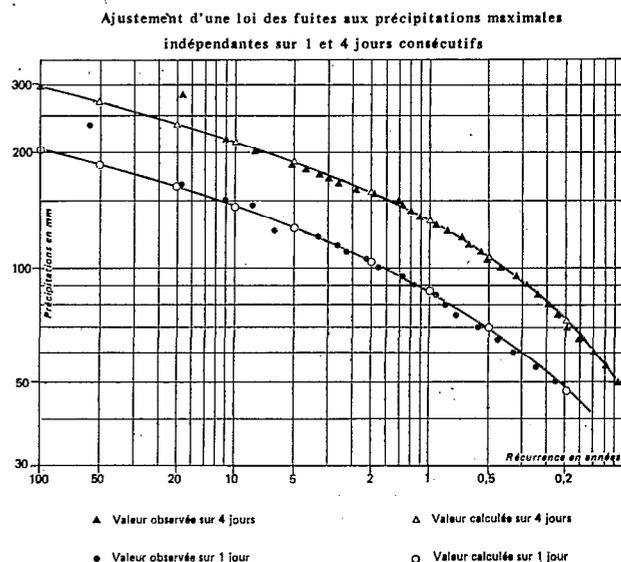


Fig. 7

Ces résultats nous montrent que l'additivité de la loi n'est pas tout à fait respectée ; théoriquement, si les précipitations sur un jour suivent une loi des fuites de paramètres λ et s , les précipitations sur m jours doivent suivre une loi des fuites de paramètres $m\lambda$ et s . Ce résultat est à peu près vérifié lorsqu'on passe de 1 à 2 jours consécutifs et lorsqu'on passe de 3 à 4 jours consécutifs. De nouveau, on remarque deux populations différentes : précipitations sur 1 ou 2 jours et précipitations sur 3 ou 4 jours.

On notera que la loi des fuites est très sensible à l'hétérogénéité de l'échantillon étudié, ce que nous n'avons pas observé sur les lois de Galton ou de Goodrich. Un plus grand nombre d'applications nous permettrait de savoir si il y a une bonne corrélation entre pas de temps minimum pour que la loi des fuites s'ajuste bien et soit additive et pas à partir duquel les précipitations peuvent être considérées comme indépendantes. Il faudrait en particulier faire ces applications sur des précipitations caractéristiques d'autres types de climat que la Guadeloupe.

Une fois le choix des seuils de troncature effectué, l'ajustement donné par la loi des fuites est comparable aux ajustements donnés par les lois de Goodrich et Galton. Un test du χ^2 effectué sur les ajustements obtenus pour les précipitations maximales journalières ne nous a pas permis de définir le meilleur ajustement.

Pour illustrer les résultats obtenus, nous avons reporté dans le tableau ci-après les périodes de retour observées et théoriques obtenues pour les 6 premières valeurs de l'échantillon des précipitations sur 2 jours, puis sur 4 jours consécutifs.

Nombre de dépassement	Valeur (mm)	Période de retour (années) sur précipitations de 2 jours			
		Observée	Calculée par loi des Fuites	Calculée par loi de Goodrich	Calculée par loi de Galton
1	270	56,0	170,1	94,5	72,1
2	215	18,7	29,8	22,0	20,3
3	185	11,2	11,5	9,7	9,5
4	170	8,0	7,4	6,4	6,4
5	165	6,2	6,1	5,5	5,6
6	155	5,1	4,5	4,1	4,2

Période de retour (années) sur précipitations de 4 jours					
2	280	18,7	70,2	44,8	27,3
3	215	11,2	10,8	8,7	7,4
4	200	8,0	7,0	5,9	5,3
6	185	5,1	4,6	4,0	3,8
7	180	4,3	4,0	3,5	3,4
8	175	3,7	3,4	3,1	3,0

Il faut souligner que l'échantillon contenant peu d'années d'observation, 28 années de précipitations journalières, la période de retour observée pour la première valeur n'est pas très significative. On peut quand même remarquer que la loi des fuites lui donne une probabilité d'apparition plus faible que les autres lois. Pour les autres valeurs, c'est la loi des fuites qui se rapproche le plus des valeurs observées.

Sur la figure 7, nous avons représenté, pour les précipitations maximales indépendantes de 1 et 4 jours consécutifs, les valeurs observées et l'ajustement donné par la loi des fuites. L'ajustement est satisfaisant pour toutes les valeurs, sauf pour la valeur maximale dans les 2 cas ; nous avons souligné que, du fait de la petite taille de l'échantillon, cette valeur pouvait avoir une fréquence plus faible que celle observée.

CONCLUSION

Malgré une formulation mathématique peu simple, la loi des fuites s'avère d'un usage très facile parce qu'elle ne nécessite pas de tronquage en fréquence. Elle est particulièrement recommandée pour tout ajustement sur un échantillon présentant des valeurs nulles. Dans les applications, nous avons pu voir qu'une simple étude de l'efficacité des différentes méthodes (voir figure 3), nous permettait de choisir la meilleure méthode d'estimation des paramètres λ et s , sans avoir à utiliser un ordinateur.

Pour des précipitations journalières avec troncature, nous avons montré que la loi des fuites donnait d'aussi bons ajustements que les lois de Galton ou Goodrich, dans le cas des données relevées à la Guadeloupe. Cette dernière application nous a aussi montré que la loi des fuites était très sensible à l'hétérogénéité, ou plutôt la non-indépendance des valeurs d'un échantillon. En appliquant la loi des fuites à un grand nombre d'échantillons différents de précipitations, il serait intéressant d'étudier les relations entre types de précipitations, seuil minimum d'indépendance des événements et qualité de l'ajustement de la loi des fuites.

6. BIBLIOGRAPHIE

- ABRAMOWITZ (M.), STEGUN (I. A.) – Handbook of Mathematical Functions. Dover Publications, New-York.
- BABUSIAUX (C.) – 1969 – Etude statistique de la loi des fuites. Thèse 3^{ème} cycle, Faculté des sciences de Paris.
- BERNIER (J.) – 1967 – Sur la théorie du renouvellement et son application en hydrologie. E.D.F., Laboratoire national d'hydraulique, Chatou, HYD. 67/10.
- BERNIER (J.), FANDEUX (D.) – 1967 – Théorie du renouvellement. E.D.F., Laboratoire national d'hydraulique, Chatou, HYD. 67/35.
- BRUNET-MORET (Y.) – 1978 – La loi des fuites. Note provisoire, ORSTOM.
- BRUNET-MORET (Y.) – 1969 – Etude de quelques lois utilisées en hydrologie. Cah. ORSTOM, série Hydrol., vol. VI, n° 3.
- BRUNET-MORET (Y.) – 1974 – Distribution exponentielle généralisée. Cah. ORSTOM, série Hydrol., vol. XI, n° 4.
- EINSTEIN (H. A.) – The application of radioactive tracers in the study of sands movements in rivers. Institute for development of water resources, Ioroslav Cerni, Beograd.
- MORLAT – La distribution des débits de fuite sur une conduite de gaz. E.D.F., Etudes économiques générales n° 1021.
- WILKS (S.) – 1962 – Mathematical statistics, Princeton University Press.

ANNEXE 1

FONCTIONS DE BESSEL

La fonction de Bessel, modifiée, d'ordre n , s'écrit :

$$I_n(Z) = \sum_{i=0}^{\infty} \frac{(Z/2)^{2i+n}}{i! \Gamma(i+n+1)}$$

Γ : fonction Eulérienne de deuxième espèce = fonction gamma :

$$\Gamma(x) = \int_0^{\infty} u^{x-1} \cdot e^{-u} \cdot du$$

Nous utiliserons surtout les fonctions de Bessel modifiées d'ordre 0 et 1.

ordre 0
$$I_0(Z) = 1 + \frac{(Z/2)^2}{1!1!} + \dots + \frac{(Z/2)^{2i}}{i!i!} + \dots$$

ordre 1
$$I_1(Z) = \frac{(Z/2)^1}{0!1!} + \frac{(Z/2)^3}{1!2!} + \dots + \frac{(Z/2)^{2i+1}}{i!(i+1)!} + \dots$$

Comme principale propriété utilisée, il y a la formule de récurrence suivante :

$$I'_n(Z) = I_{n+1}(Z) + \frac{n}{Z} \cdot I_n(Z) = I_{n-1}(Z) - \frac{n}{Z} \cdot I_n(Z)$$

soit en particulier :
$$I'_1(Z) = I_0(Z) - \frac{1}{Z} \cdot I_1(Z)$$

Approximation des fonctions de Bessel :

$$t = x/3,75$$

$$- 0 \leq x \leq 3,75$$

$$I_0(x) = 1 + 3,51562 \cdot t^2 + 1,2067492 \cdot t^4 + 3,0899424 \cdot t^6 + 0,2659732 \cdot t^8 + 0,00301532 \cdot t^{10} + 0,00032411 \cdot t^{12} + \varepsilon \quad |\varepsilon| < 1,6 \cdot 10^{-7}$$

$$I_1(x)/x = 0,5 + 0,87890594 \cdot t^2 + 0,51498869 \cdot t^4 + 0,15084934 \cdot t^6 + 0,2658733 \cdot t^8 + 0,00301532 \cdot t^{10} + 0,00032411 \cdot t^{12} + \varepsilon \quad |\varepsilon| < 8 \cdot 10^{-9}$$

$$- 3,75 < x < \infty$$

$$\sqrt{x} \cdot e^{-x} \cdot I_0(x) = 0,39894228 + 0,01328592 \cdot t^{-1} + 0,00225319 \cdot t^{-2} - 0,00157565 \cdot t^{-3} + 0,00916281 \cdot t^{-4} - 0,02057706 \cdot t^{-5} + 0,02635537 \cdot t^{-6} - 0,01647633 \cdot t^{-7} + 0,00392377 \cdot t^{-8} + \varepsilon \quad |\varepsilon| < 1,9 \cdot 10^{-7}$$

$$\sqrt{x} \cdot e^{-x} \cdot I_1(x) = 0,39894228 - 0,03988024 \cdot t^{-1} - 0,00362018 \cdot t^{-2} + 0,00163801 \cdot t^{-3} - 0,01031555 \cdot t^{-4} + 0,02282967 \cdot t^{-5} - 0,02895312 \cdot t^{-6} + 0,01787654 \cdot t^{-7} - 0,00420059 \cdot t^{-8} + \varepsilon \quad |\varepsilon| < 2,2 \cdot 10^{-7}$$

ANNEXE 2

SOUS-PROGRAMMES FORTRAN

1. SIMULATION D'UNE LOI DES FUITES
2. FONCTION DE DISTRIBUTION
3. INVERSION DE LA FONCTION DE DISTRIBUTION
4. CALCUL DES PARAMÈTRES, ÉCHANTILLON SANS TRONCATURE
5. CALCUL DES PARAMÈTRES, ÉCHANTILLON AVEC TRONCATURE

- 1 -

```

DIMENSION FL(200),X(400)
COMMON XK1,XK2
READ(5,20)IX
READ(5,22)N
READ(5,21)XLAM,S,XH
20 FORMAT(1B)
21 FORMAT(3F8.4)
22 FORMAT(I5)
WRITE(6,40)N,XLAM,S,XH
40 FORMAT(' LOI DES FUITES SIMULATION DE',I4,' VALEURS',/, ' PARAM.
1 DE FORME ',F5.2,' PARAM. D ECHELLE ',F5.2,' TRONCATURE ',F6.3,/)
WRITE(6,98)IX
98 FORMAT(5X, ' VALEUR DE DEPART ',I8,/)
Q=EXP(-XLAM)
FL(1)=Q
R=1.0
I=1
12 I=I+1
R=R*XLAM/(I-1)
FL(I)=FL(I-1)+Q*R
IF(FL(I).LT.0.9999)GO TO 12
M=I
J=0
N1=0
VA=0.
XM=0.
8 J=J+1
IX=IABS(IX*65539)
F=IX/2147483648.DO
DO 10 L=1,M
IF(F.LT.FL(L)) GO TO 9
10 CONTINUE
9 L=L-1
X(J)=0.
IF(L.EQ.0)GO TO 6
DO 7 I=1,L
IX=IABS(IX*65539)
7 X(J)=X(J)+DABS(C*DLOG((2147483648.DO/IX)/S))
6 IF(X(J).GE.XH)GO TO 5
J=J-1
N1=N1+1
GO TO 8
5 XM=XM+X(J)
VA=VA+X(J)*X(J)
IF(J.LT.N)GO TO 8
XX=XM/(N+N1)
XM=XM/N
VX=(VA-XX*XX*(N+N1))/(N+N1-1)
VA=(VA-XX*XM*N)/(N-1)
WRITE(6,41)XM,VA,N1
41 FORMAT(15X, ' MOYENNE =',F9.4,' VARIANCE =',F9.4,' NB DE VALEURS
2<SEUIL ',I5,/)
S=VX/(2*XX)
XLAM=2*XX**2/VX
WRITE(6,31)XLAM,S
31 FORMAT(10X, 'ESTIMATION DES PARAMETRES PAR LA METHODE DES MOMENTS '
1,/, 5X, 'PARA. DE FORME LAMBDA =',F8.4,' D ECHELLE S:',F8.4)
2,/)
K=N/10
DO 2 I=1,K
KI=(I-1)*10
2 WRITE(6,42)(X(KI+J),J=1,10)
42 FORMAT(/, 5X, 10F8.4,/)
CALL OPTIM(XLAM,S,N,N1,X,XH)
WRITE(6,32)XLAM,XK1
32 FORMAT(/, 10X, 'ESTIMATION DES PARAMETRES DE LA LOI TRONQUEE PAR LE
1 MAXIMUM DE VRAISEMBLANCE ',/, 22X, 'PARAMETRE DE FORME LAMBDA =
2',E9.6,' +/- 1',E9.6)
WRITE(6,33)S,XK2
33 FORMAT(/, 34X, 'D ECHELLE S =',F9.6,' +/- 1',F9.6)
STOP
END

```

— 2 —

```

FUNCTION FFUITE(Y, XLAM, S)
REAL*8 A, B, C, D, E, F
F=DEXP(XLAM/2. DO)
IF(Y. NE. 0) GO TO 11
FFUITE=DEXP(-XLAM*1. DO)
GO TO 13
11 A=F-1. DO/F
B=DEXP(-Y/2. DO/S)
C=1. DO
D=A*B
E=0. DO
12 E=E+1. DO
B=B*Y/S/E
C=C*XLAM/E
A=A-C/F
D=D+A*B
IF(A*B/D. GT. 1. D-6) GO TO 12
FFUITE=1. DO-D/F/DEXP(Y/2. DO/S)
13 RETURN
END

```

```

FUNCTION VFUITE(P, XLAM, S)
REAL*8 F
A=XLAM*S
B=0.
C=0.
K=4
IC=0
21 F=FFUITE(A, XLAM, S)*1. DO
IC=IC+1
IF(IC. LT. 100) GO TO 25
WRITE(6, 60)
60 FORMAT(/, 70X, 'CONVERGENCE NON ATTEINTE ')
GO TO 22
25 CONTINUE
IF(DABS((F-P)/P). LT. 1. D-6) GO TO 22
IF(F. LT. P) GO TO 23
B=A
A=(A+C)/2
K=1
GO TO 21
23 C=A
A=(K*A+B)/2
GO TO 21
22 CONTINUE
VFUITE=A
RETURN
END

```

— 4 —

```

SUBROUTINE NTFUIT(XLAM, S, X, N, XM)
DIMENSION X(N), H(2)
REAL*8 A, B
H(1)=0. 01
XLAM=140.
SI=-1.
1 H((3-IFIX(SIGN(1., SI)))/2)=XLAM
XLAM=(256. +H(1)+H(2)-256. )/2.
IF((XLAM-H(1))*(XLAM-H(2)). GE. 0. ) GO TO 4
A=0. DO
DO 3 I=1, N
IF(X(I). EQ. 0. ) GO TO 3
B=XLAM/1. 875DO*SQRT(X(I)/XM)
IF(B. GT. 1. DO) GO TO 2
B=B*B
A=A+(((100. /B+351. 56229)/B+308. 994)/B+120. 67492+B*(26. 59732+B*(3. 6
10768+B*. 45813)))/((50. /B+87. 870594)/B+51. 498869)/B+15. 084934+D*(2
2. 658733+B*(. 301532+B*. 032411)))
GO TO 3
2 A=A+B*(((B*39. 894228+1. 328592)*B+. 225319)*B-. 157565)*B+. 916281+((
1(. 392377/B-1. 647633)/B+2. 635537)/B-2. 057706)/B)/(((B*39. 894228-3.
298802)*B-. 362018)*B+. 163801)*B-1. 031555+(((1. 787654-. 420059/B)/B-2
3. 895312)/B+2. 282967)/B)*3. 75DO
3 CONTINUE
SI=A-2. *N*XLAM
IF(AINT(SI*5. E5). NE. 0. ) GO TO 1
4 S=XM/XLAM
RETURN
END

```

— 5 —

```

SUBROUTINE OPTIM(XLAM, S, N, N1, X, XH)
DIMENSION X(N)
COMMON XK1, XK2
REAL*8 V, XX
MM=0
XX=-1. E75
XK1=2. 5
XK2=2. 5
XL=XLAM
XS=S
3 PE=S-2*XS/XK2
II=2
JJ=2
DO 1 I=1, 3
PE=PE+XS/XK2
PF=XLAM-2*XL/XK1
DO 1 J=1, 3
PF=PF+XL/XK1
CALL VRAIS(PF, PE, N, N1, X, XH, V)
IF(V. LE. XX) GO TO 1
JJ=J
II=I
FF=PF
EE=PE
XX=V
1 CONTINUE
MM=MM+1
IF(MM. GT. 200) GO TO 4
IF(II*JJ. NE. 4) GO TO 2
XK1=XK1*2
XK2=XK2*2
7 IF(XK1. GT. 1. E+08. OR. XK2. GT. 1. E+08) GO TO 4
2 XLAM=FF
S=EE
GO TO 3
4 WRITE(6, 12)MM
12 FORMAT(/, 10X, ' NOMBRE D ITERATION ', I3, /)
RETURN
END

SUBROUTINE VRAIS(XLAM, S, N, N1, X, XH, V)
DIMENSION X(N)
REAL*8 A, B, C, E
REAL*8 V
V=0.
DO 1 I=1, N
A=1. DO
C=1. DO
E=0. DO
B=XLAM*X(I)/S*1. DO
12 E=E+1. DO
C=C*B/(E*(E+1. DO))
A=A+C
IF(C/A. GT. 1. D-6) GO TO 12
1 V=V+DLOG(A)-X(I)/S
V=V-N*XLAM+N*DLOG(XLAM/S*1. DO)
FC=FFUITE(XH, XLAM, S)
V=V+N1*DLOG(FC*1. DO)
RETURN
END

```