Application de lois tronquées aux distributions de précipitations journalières

Hélène Lubès [a]

Résumé

L'objet de cet article est de poser des questions relatives à l'utilisation de méthodes d'ajustement de lois de probabilité tronquées sur des échantillons de données hydrologiques.

Il souligne la nécessité de définir précisément le problème à résoudre avant de mettre en oeuvre une méthode statistique de ce type, dont l'adéquation au problème pourrait n'être qu'apparente.

Pour illustrer ce propos, nous reprenons ici l'exposé de la méthode développée par Brunet-Moret (1973) qui est à l'origine de cette réflexion, en considérant son application au cas de précipitations journalières mesurées par un pluviomètre.

[a] Laboratoire d'hydrologie Centre ORSTOM de Montpellier, B.P.5045, 34032 Montpellier Cedex 1

1 Lois tronquées et échantillons censurés : définitions théoriques

1.1 Tronquage et Censure

Il nous semble nécessaire de commencer cet article par quelques définitions (Kendall et Stuart, 1977).

Soit la variable X, distance d'un impact de tir au centre d'une cible circulaire de rayon R.

On ne peut observer la variable X que pour des tirs touchant la cible.

Si on ne connaît pas le nombre de tirs effectués, on ne peut retenir que les m valeurs de X observées sur la cible, distribuées entre 0 et R.

On dit que *la distribution de X est tronquée à droite.*

Si on connaît le nombre n de tirs dirigés vers la cible, et le nombre m de projectiles l'ayant atteinte, on en déduit que $n - m = r$ autres valeurs de X supérieures à R existent. En d'autres termes, on observe les m premières valeurs de la série statistique $x_1, x_2 \ldots x_m$ dans un échantillon de taille n.

L'échantillon des valeurs de la variable X est appelé échantillon censuré à droite en R.

On déduit de cet exemple que la censure caractérise l'échantillon alors que le "tronquage" est une propriété de la distribution de probabilité.

Lois tronquées et échantillons censurés peuvent être définis de la même façon inférieurement (ou à gauche), ou supérieurement (ou à droite), ou supérieurement et inférieurement (ou à droite et à gauche).

1.2 Lois de probabilités tronquées

Considérons la loi de probabilité d'une variable aléatoire X variant de $-\infty$ à $+\infty$. $f(x)$ désigne la fonction densité de probabilité, et $F(x)$ la fonction de répartition,

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

On suppose que les n valeurs observées de cette variable n'appartiennent qu'à une partie $[a, b]$ de son domaine de variation. La probabilité que la variable prenne une valeur à l'extérieur du domaine $[a, b]$ est non nulle mais inconnue. On peut envisager de n'étudier la distribution de la variable aléatoire que sur la partie $[a, b]$. On définit ainsi la loi de probabilité tronquée de fonction de répartition $\xi(x)$ telle que :

$$\xi(x) = \Pr_{a \leq x \leq b}(X \leq x) = \frac{F(x) - F(a)}{F(b) - F(a)}$$

$$\xi(a) = 0 \text{ et } \xi(b) = 1$$

et de fonction densité de probabilité :

$$\frac{\delta\xi(x)}{\delta x} = \frac{f(x)}{\int_a^b f(x)dx}$$

Si on fait choix d'une expression analytique fonction de paramètres θ_k pour la loi de distribution $f(x)$, une estimation des paramètres θ_k à partir des seules n

valeurs observées disponibles, appartenant à l'intervalle $[a, b]$ peut être obtenue en utilisant la méthode du maximum de vraisemblance.

La méthode du maximum de vraisemblance consiste à trouver les paramètres θ_k qui rendent maximale la probabilité d'obtenir l'échantillon des valeurs observées avec la loi envisagée.

La fonction de vraisemblance à maximiser $L(x, \theta)$ s'écrit donc ici :

$$L(x, \theta) = \prod_{i=1}^{n} \frac{\delta\xi(x_i)}{\delta x}$$

$$L(x, \theta) = \frac{\prod_{i=1}^{n} f(x_i)}{\left[\int_a^b f(x)dx \right]^n}$$

1.3 Echantillons censurés

On définit deux types de censures.

La censure de type I désigne des valeurs censurées plus petites ou égales et/ou plus grandes ou égales à une ou plusieurs valeurs prédéterminées. Les points de censure sont connus et fixes, et le nombre d'observations censurées est une variable aléatoire.

Lorsqu'une proportion fixée d'un échantillon de taille n est censurée, la censure est dite de type II. Les nombres d'observations plus petites ou plus grandes que les points de censure sont fixés a priori, tandis que les points de censure sont des variables aléatoires.

D'un point de vue pratique, la censure de type II se rencontre fréquemment lorsque la variable X est le temps. Il peut s'agir par exemple de la période de défaillance d'une pièce soumise à des tests, le temps expérimental étant limité. L'expérience est arrêtée quand, parmi n observations, les m premières sont disponibles. Par suite, la censure de type II est généralement à droite.

Nous ne nous intéresserons ici qu'aux censures de type I rencontrées en hydrologie.

Soit donc un échantillon censuré de type I.

On étudie la variable aléatoire X dont on connaît sur n réalisations, m valeurs observées dans un domaine $[a, b]$:

$$n = m + r_1 + r_2$$

r_1 désigne le nombre de réalisations inférieures ou égales à a, et r_2 le nombre de réalisations supérieures ou égales à b.

L'estimation des paramètres de la loi de probabilité théorique $f(x)$ qui est supposée décrire la population de la variable X, doit tenir compte des données censurées.

En effet, si on estime classiquement la moyenne et l'écart-type de la population à partir des m données observées sans tenir compte des données censurées, on aura pour une censure à gauche par exemple, une surestimation de la moyenne, et une sous-estimation de l'écart-type.

La méthode du maximum de vraisemblance permet de tenir compte du nombre de données censurées.

La fonction de vraisemblance à maximiser est la suivante :

$$L(x,\theta) = \left[\ \prod_{i=1}^{n} f(x_i) \ \right] \left[\ \int_{-\infty}^{a} f(x)dx \ \right]^{r_1} \left[\ \int_{b}^{+\infty} f(x)dx \ \right]^{r_2}$$

$$L(x,\theta) = \left[\ \prod_{i=1}^{n} f(x_i) \ \right] \left[\ F(a) \ \right]^{r_1} \left[\ 1 - F(b) \ \right]^{r_2}$$

1.4 Echantillon censuré et loi tronquée : applications pratiques

Les phénomènes de censure sont fréquents en hydrologie, et d'une manière plus générale dans les disciplines qui traitent de l'environnement.

Les censures à gauche concernent, par exemple, les valeurs inférieures à un seuil de sensibilité d'un appareil de mesure (Gilliom *et al.* 1986).

Les censures à droite sont relatives aux données historiques, ou aux hauteurs pluviométriques dépassant la capacité du pluviomètre (Hosking *et al.* 1986).

Les données censurées ne présentent pas un gros problème d'interprétation, si en premier lieu on s'intéresse aux valeurs qui sont supérieures à un point de censure à gauche, ou inférieures à un point de censure à droite.

Mais il est parfois nécessaire d'estimer les probabilités d'occurrence de valeurs appartenant au domaine censuré. Ainsi quelques critères chimiques de qualité des eaux sont définis en-deçà des limites de détection des procédés d'analyse de laboratoire (Gilliom *et al.* 1986).

Depuis une trentaine d'années, des techniques propres à l'analyse de fréquence de données censurées ont été développées pour répondre à ce problème.

Plusieurs méthodes d'estimation des paramètres de distributions de probabilité relatives à des ensembles de données censurées ont été mises au point à partir des seules observations non censurées, c'est-à-dire réellement observées de l'échantillon.

Ces méthodes sont généralement spécifiques à la loi de probabilité qui est supposée représenter la distribution de la variable que l'on ne connaît expérimentalement qu'au travers d'échantillons censurés. Les lois normale, log-normale, Gamma, Gumbel et GEV (Generalized Extreme Value distribution) ont donné lieu à de tels développements, et sans être abondante, la littérature statistique fait état de ces recherches (Cohen 1959 - Gilliom *et al.* 1986, Harter 1967, Leese 1973, Phien *et al.* 1989)

Les questions statistiques qui se posent en hydrologie et sont traitées par la théorie des lois tronquées sont plus complexes. Nous avons choisi de présenter ici un problème de "tronquage" traité par Brunet-Moret (1969, 1973, 1975), et appliqué aux distributions de précipitations journalières.

2 Les distributions de précipitations journalières. Application de la théorie des lois tronquées.

La méthode proposée par Brunet-Moret trouve sa justification dans l'énoncé du problème suivant.

On étudie des observations dont la borne inférieure est x_0. La fréquence des observations inférieures ou égales à x_0 n'est pas nulle alors que la densité de la loi de probabilité a priori choisie pour procéder à un ajustement statistique sur ces données est nulle ou indéfinie en ce point.

De plus, entre x_0 et x_h (x_h seuil supérieur à la borne x_0), le nombre des observations qui auraient dû être faites est inconnu, ainsi que les valeurs individuelles de ces observations.

Le seuil x_h peut être déterminé par la sensibilité de l'appareil de mesure, ou ses conditions d'emploi.

Ce cas est illustré par l'étude des hauteurs journalières de pluie mesurées par un pluviomètre, la borne x_0 étant égale à 0.

L'étude statistique de ces données est en effet confrontée aux problèmes suivants :

- La confiance accordée aux valeurs nulles est très faible puisque les quantités inférieures à 0.1 mm ne sont pas mesurées. Toutes les valeurs nulles ne représentent donc pas des jours sans pluie.

- Si l'évaporation dans le pluviomètre n'est pas négligeable, un seuil supérieur à celui de 0.1 mm défini par la sensibilité de l'appareil de mesure, doit être retenu (5 à 10 mm selon Brunet-Moret) pour partager les observations en deux groupes : les douteuses c'est-à-dire celles qui sont entachées d'une très grande incertitude relative en-deçà du seuil, et les bonnes au-delà.

On considère donc que toutes les observations inférieures à un seuil x_h strictement positif traduisent des précipitations inférieures à x_h, mais sans aucune autre précision. Les valeurs observées supérieures à x_h sont quant à elles admises comme étant de bonnes mesures, du moins jusqu'à un certain seuil supérieur au-delà duquel le pluviomètre déborde, mais qui ne fait pas l'objet de notre propos.

Parmi les précipitations inférieures à x_h, on ne connaît pas le nombre de jours où on aurait dû observer une hauteur de précipitation non nulle .

Ce constat étant fait, que cherche-t-on à étudier ?

En fait l'objectif n'est pas clairement défini, mais la méthode développée par Brunet-Moret laisse à penser qu'on s'intéresse aux probabilités d'occurrence des précipitations inférieures à x_h, et en particulier à la probabilité P_0 non nulle d'absence de pluie. En effet, la démarche est la suivante.

Le seuil x_h est fixé a priori.

Les valeurs non nulles (inférieures ou supérieures à x_h) sont distribuées selon une loi dont la fonction de répartition "à l'infini" est $1 - P_0$.

On choisit une loi de distribution de ces valeurs non nulles parmi plusieurs lois possibles dont on cherche à estimer les paramètres.

Le cas traité par Brunet-Moret est celui où la distribution des précipitations est modélisée par des lois de probabilité asymétriques à deux paramètres ($x_0 = 0$) qui ont une probabilité nulle ou indéfinie en 0.

La distinction entre valeurs nulles et jours réels sans pluie ne pouvant être faite, le nombre de valeurs nulles vraies inférieures à x_h est inconnu, de même par conséquent que le nombre de valeurs non nulles vraies. La distribution des valeurs non nulles est donc tronquée à gauche en x_h.

Par conséquent, si on veut déterminer la distribution théorique de l'ensemble des valeurs non nulles, inférieures ou supérieures à x_h, on choisit une loi de densité de probabilité $g(x)$ fonction de k (ici $k = 2$) paramètres (le choix de cette loi n'est pas traité ici).

Pour obtenir une estimation de ces paramètres à partir des n valeurs strictement supérieures à x_h, on maximise la fonction de vraisemblance suivante :

$$L(x > 0, \theta_k) = \frac{\prod_{i=1}^{n} g(x_i)}{\left[\int_{x_h}^{+\infty} g(x)dx \right]^n}$$

Si $G(x)$ désigne la fonction de répartition de la loi de probabilité des valeurs non nulles, et $F(x)$ la fonction de répartition de la loi de probabilité des données de pluie nulles ou non nulles, la relation existant entre ces deux fonctions est :

$$G(x) = \frac{F(x) - P_0}{1 - P_0}$$

ou

$$F(x) = (1 - P_0)G(x) + P_0 \tag{1}$$

Par ailleurs, le nombre total N d'observations nulles ou non nulles étant supposé connu, le rapport n/N désigne la proportion de valeurs strictement supérieures à x_h. On peut donc écrire :

$$1 - F(x_h) \approx \frac{n}{N} \tag{2}$$

De ces deux dernières relations, Brunet-Moret (1973) déduit une estimation de P_0.

En effet la relation (1) appliquée en x_h donne

$$F(x_h) = (1 - P_0)G(x_h) + P_0$$

ce qui dans la relation (2) conduit à :

$$1 - (1 - P_0)G(x_h) - P_0 \approx \frac{n}{N}$$

d'où

$$P_0 \approx 1 - \frac{n}{N(1 - G(x_h))} \tag{3}$$

Brunet-Moret (1973) inclut dans le nombre n, les n_h valeurs égales à x_h. Si n_h n'est pas petit devant N l'approximation représentée par la relation (2) n'est plus acceptable. C'est la raison pour laquelle nous préférons ne retenir que les valeurs strictement supérieures à x_h.

A partir de P_0 et des paramètres de la fonction $G(x)$, la relation (1) définit la distribution de fréquence des hauteurs journalières de pluie mesurées par un pluviomètre.

L'estimation des paramètres de la loi des valeurs non nulles, et de P_0, repose sur le choix du seuil x_h. La valeur de ce seuil est choisie a priori en raison des considérations "physico-expérimentales" présentées précédemment. Mais sous réserve que l'information supérieure à x_h représente un pourcentage significatif de l'information totale, la démarche présentée reste théoriquement valable.

Lors des applications qui sont faites de cette méthode le seuil x_h est rarement connu expérimentalement, et entre 0.1 et 10 mm (et peut être au-delà...) le choix est difficile. Aussi la valeur du seuil x_h est le plus souvent déterminée par tâtonnements de façon à ce que l'ajustement de la loi de probabilité choisie sur l'ensemble des valeurs supérieures à x_h soit le plus satisfaisant possible.

Cette utilisation est-elle valide au regard des objectifs fixés, et en particulier de l'estimation de P_0 ?

Nous avons cherché à mettre en oeuvre cette méthode de cette façon sur un cas concret. L'application et l'interprétation des résultats font l'objet du paragraphe suivant.

3 Application à l'étude des précipitations journalières d'Abidjan-ville

L'échantillon disponible est constitué de N = 1092 valeurs de données de pluie journalières d'Abidjan-Ville de 1966 à 1977, au cours des mois de septembre, octobre et novembre, deuxième saison humide dans l'année. Nous avons choisi a priori d'ajuster une loi log-normale à deux paramètres pour décrire la distribution des valeurs non nulles tronquée en x_h.

La fonction densité de probabilité est :

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{\left[-\frac{(Log(x)-Log(s))^2}{2\sigma^2}\right]}$$

$$s > 0, \sigma > 0$$

Soit n le nombre de valeurs strictement supérieures à x_h.

La fonction de vraisemblance à maximiser s'écrit :

$$L(s,\sigma) = \frac{\prod_{i=1}^{n} g(x_i)}{\left[\int_{x_h}^{+\infty} g(x)dx\right]^n}$$

Le logarithme de la fonction de vraisemblance est donc égal à :

$$Log(L(s,\sigma)) = \sum_{i=1}^{n} Log(g(x_i)) - nLog(1 - G(x_h))$$

G désigne la fonction de répartition de la loi log-normale.

Les développements de la méthode de résolution pour l'obtention des paramètres s et σ sont rassemblés en annexe.

Nous avons mis en oeuvre cette méthode pour différentes valeurs de x_h allant de 0.09mm à 14.09mm par pas de 2 mm. Dans chaque cas nous avons estimé la probabilité P_0 donnée par la relation (3). Les effectifs des échantillons traités sont les suivants :

seuil (mm)	effectif (observations)
0.09	308
2.09	220
4.09	178
6.09	149
8.09	124
10.09	108
12.09	95
14.09	83

Les graphiques (fig. 1 à fig. 3) montrent l'influence très sensible du seuil pour des valeurs inférieures à 10.09 mm, sur l'estimation de s, σ et P_0.

Pour juger de la qualité des ajustements, nous avons porté sur un graphique gausso-logarithmique, d'une part les valeurs observées supérieures au seuil x_h et leur fréquence empirique de non-dépassement calculée à partir de la formule de Hazen

$$\left(\frac{\text{rang de l'observation - 0.5}}{n}\right),$$

et d'autre part la fonction de répartition théorique.

Celle-ci a été obtenue de la façon suivante.

Nous avons déterminé en inversant la fonction de répartition $G(x)$ de la loi log-normale identifiée, les valeurs $x(p)$ de la variable X, hauteur non nulle de pluie, correspondant aux probabilités de non-dépassement $p = 0.01$, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99. Comme l'adéquation de l'ajustement ne peut être étudiée qu'à partir des seules observations supérieures à x_h, les probabilités correspondantes aux valeurs $x(p)$ ont été réestimées en ne considérant pour domaine de variation de la variable X que les valeurs supérieures à x_h.

La probabilité de non-dépassement de la valeur $x(p)$ est alors :

$$\frac{p - G(x_h)}{1 - G(x_h)}$$

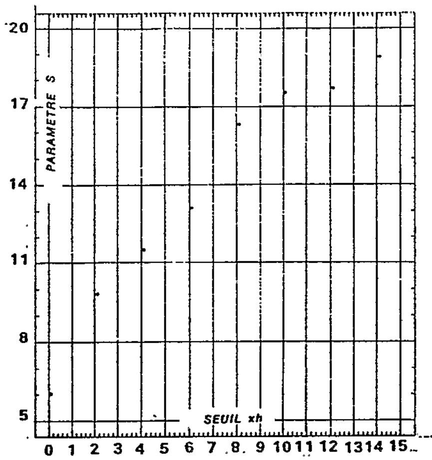
L'examen des graphiques (figure 4) montre que le seuil choisi influe sur l'ajustement de la loi théorique aux observations, jusque dans la gamme des fréquences rares de dépassement.

A partir de la valeur 8.09mm du seuil, les fréquences théoriques de dépassement inférieures à 0.16 (variable normale réduite supérieure à 1) s'écartent de plus en plus des valeurs de fréquence empirique. Si la qualité de l'ajustement est appréciée au regard de l'ensemble des fréquences, la simple discrimination graphique conduit à retenir un seuil x_h voisin de 4.09mm. De fait, il semble, sous l'hypothèse qu'un seuil x_h optimal existe, que celui-ci soit compris entre 2.09 et 6.09mm.

La probabilité d'occurrence P_0 d'un jour sans pluie, obtenue par la relation (3), est alors voisine de 0.81, ce qui conduit à 885 valeurs nulles sur un échantillon de 1092 valeurs. Si on estime que le nombre de 0 de l'échantillon (ici 784) est une surestimation du nombre de vraies valeurs nulles, quelle confiance doit-on accorder à cette estimation de P_0 ?

Ces constatations conduisent donc aux réflexions suivantes.

A priori il n'y a aucune raison évidente pour que la valeur x_h, déterminée par tâtonnement en fonction de la qualité de l'ajustement d'une loi de probabilité



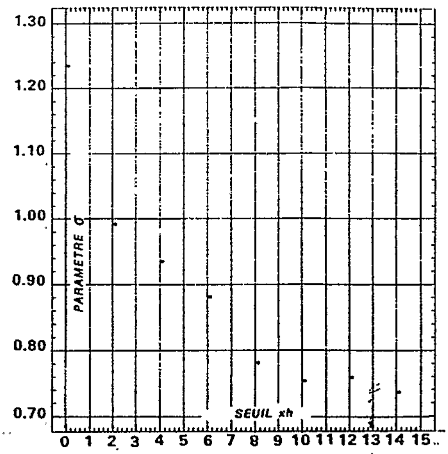


Figure 1: (Gauche) Estimation de s en fonction de x_h

Figure 2: (Droite) Estimation de σ en fonction de x_h

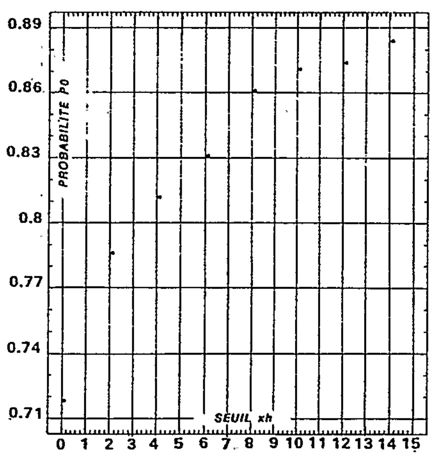


Figure 3: Estimation de P_0 en fonction de x_h

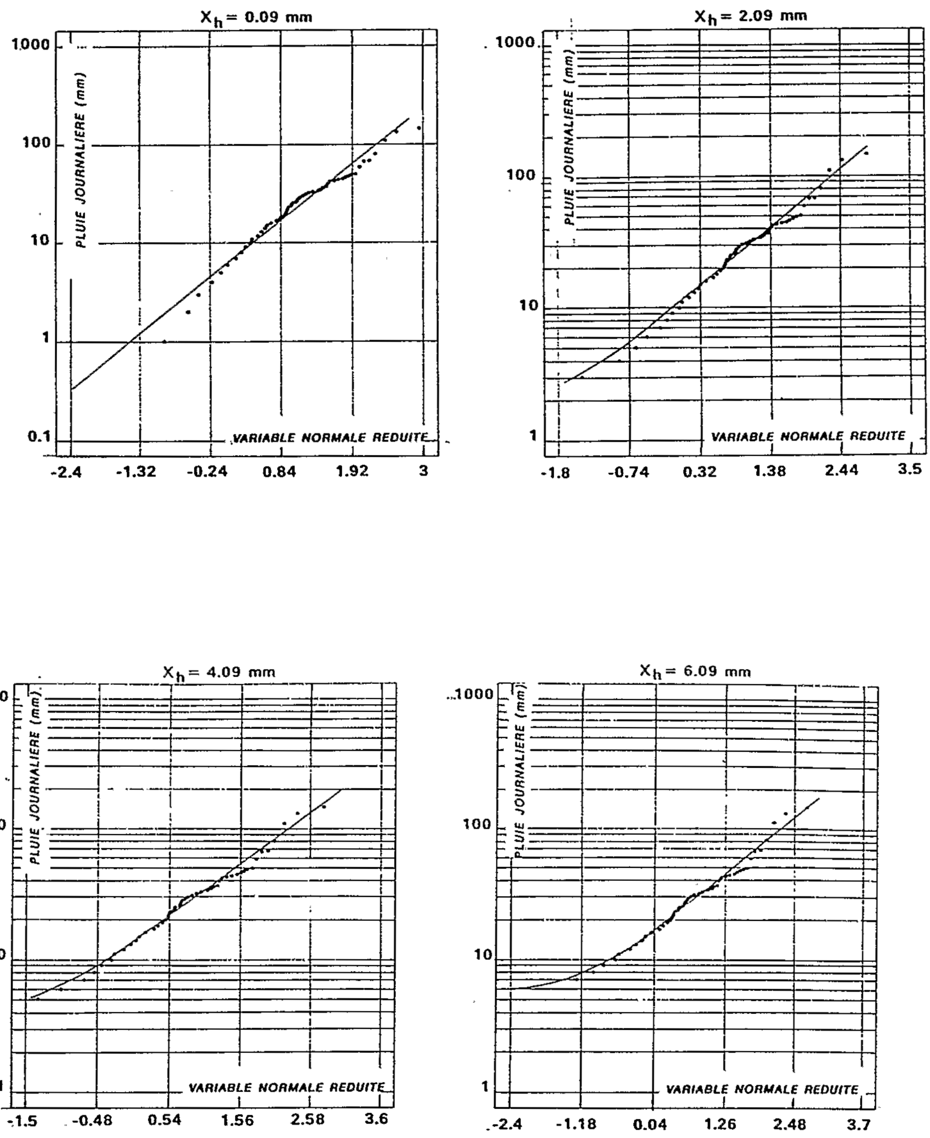


Figure 4 (début) Ajustement d'une loi log-normale ; Précipitations journalières d'Abidjan-Ville

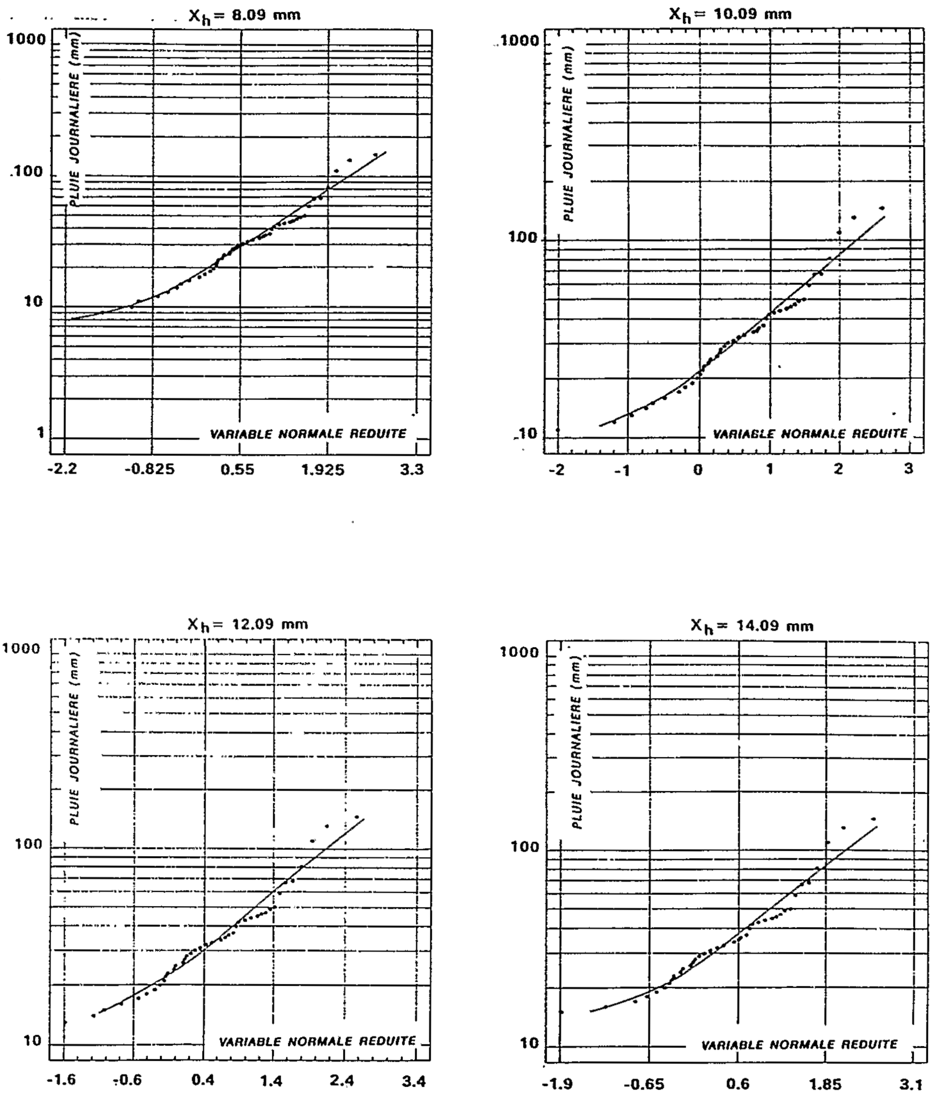


Figure 4 (fin) Ajustement d'une loi log-normale ; Précipitations journalières d'Abidjan-Ville théorique réalisé sur les données supérieures à x_h, ait une relation avec le seuil physico-expérimental qui partage les observations en deux groupes , les bonnes et les douteuses.

La façon qui vient d'être décrite de mettre en oeuvre la méthode ne répond-elle pas en fait à un tout autre objectif que celui d'estimer les probabilités des petites pluies et des jours sans pluie ? N'est-ce pas plutôt le moyen de rechercher un seuil qui permet un bon ajustement sur les valeurs fortes ? Mais est-il alors nécessaire de prendre en considération l'information des faibles précipitations ?

4 Conclusion

Beaucoup de questions et pas de réponses en conclusion de cet exposé.
En fait, à travers cet exemple, notre propos est de mettre en évidence :

- d'une part, même si cela apparaît comme une généralité, la nécessité de définir précisément les objectifs de l'étude statistique à réaliser, afin d'étudier l'opportunité d'appliquer une méthode donnée. S'il s'agit par exemple de procéder à un ajustement de loi de probabilité sur des valeurs fortes, l'utilisation d'une méthode aussi complexe que celle qui a été présentée, tant au niveau de ses fondements qu'à celui des développements mathématiques ou numériques auxquels elle peut conduire avec certaines lois, se justifie-t-elle ?

- d'autre part, les problèmes d'interprétation relatifs à l'application de lois tronquées. Sur quelles hypothèses repose la mise en oeuvre de lois tronquées, et dans quelle mesure les extrapolations que l'on en déduit sont-elles valides ?

5 Annexe

Estimation des paramètres s et σ de la loi log-normale
La fonction densité de probabilité est :

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{\left[-\frac{(Log(x)-Log(s))^2}{2\sigma^2}\right]}$$

$$s > 0, \sigma > 0$$

Soit n le nombre de valeurs strictement supérieures à x_h. La fonction de vraisemblance à maximiser s'écrit :

$$L(s,\sigma) = \frac{\prod_{i=1}^{n} g(x_i)}{\left[\int_{x_h}^{+\infty} g(x)dx\right]^n}$$

Le logarithme de la fonction de vraisemblance est donc égal à :

$$Log(L(s,\sigma)) = \sum_{i=1}^{n} Log(g(x_i)) - nLog(1 - G(x_h)) \tag{4}$$

G désigne la fonction de répartition de la loi log-normale. En portant l'expression de $g(x)$ dans la relation (4), il vient :

$$Log(L(s,\sigma)) = -nLog(\sqrt{2\pi}) - nLog(\sigma) - \sum_{i=1}^{n} Log(x_i)$$

$$-\frac{1}{2}\sum_{i=1}^{n} \frac{(Log(x_i) - Log(s))^2}{\sigma^2} - nLog(1 - G(x_h))$$

En annulant les dérivées de $Log(L(s,\sigma))$ par rapport à chacun des paramètres s et σ, on obtient le système d'équations :

$$\begin{cases} -\frac{1}{\sigma^2}\left[\frac{Log(s)}{s} - \frac{1}{ns}\sum_{i=1}^{n} Log(x_i)\right] + \frac{1}{1-G(x_h)}\frac{\delta G(x_h)}{\delta s} & = 0 \\ \\ -\frac{1}{\sigma} + \frac{1}{\sigma^3}\left[\frac{1}{n}\sum_{i=1}^{n}(Log(x_i))^2 - \frac{2}{n}Log(s)\sum_{i=1}^{n} Log(x_i) + (Log(s))^2\right] & \\ +\frac{1}{1-G(x_h)}\frac{\delta G(x_h)}{\delta \sigma} & = 0 \end{cases}$$

En reprenant les notations de Brunet-Moret (1973),

$$SL = \frac{1}{n}\sum_{i=1}^{n} Log(x_i)$$

$$SM = \frac{1}{n}\sum_{i=1}^{n}(Log(x_i))^2 \tag{5}$$

et

$$G(x_h) = \int_{-\infty}^{u_h} \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2}(\frac{Log(u)}{\sigma})^2} du, \quad u = \frac{x}{s}, \, u_h = \frac{x_h}{s}$$

il vient :

$$\begin{cases} -\frac{1}{\sigma^2}\left[\ \frac{Log(s)}{s} - \frac{SL}{s}\ \right] + \frac{1}{1-G(x_h)}\frac{\delta G(x_h)}{\delta s} & = 0 \\ -\frac{1}{\sigma} + \frac{1}{\sigma^3}[SM \ - 2SL\,Log(s) + (Log(s))^2] + \frac{1}{1-G(x_h)}\frac{\delta G(x_h)}{\delta \sigma} & = 0 \end{cases}$$

Sachant que :

$$\begin{aligned} \frac{\delta G(x_h)}{\delta s} = & \quad \frac{\delta G(x_h)}{\delta u_h}\frac{\delta u_h}{\delta s} = & -\frac{1}{\sqrt{2\pi} s\sigma} e^{-\frac{1}{2}(\frac{Log(u_h)}{\sigma})^2} \\ \frac{\delta G(x_h)}{\delta \sigma} = & \quad \frac{\delta G(x_h)}{\delta v_h}\frac{\delta v_h}{\delta \sigma} = & -\frac{1}{\sqrt{2\pi}}\frac{Log(u_h)}{\sigma^2} e^{-\frac{1}{2}(\frac{Log(u_h)}{\sigma})^2}, \\ v_h = \frac{Log(u_h)}{\sigma} & & \end{aligned}$$

le système se réduit à :

$$\begin{cases} \frac{1}{\sigma}(SL - Log(s)) = \frac{1}{\sqrt{2\pi}(1-G(x_h))} e^{-\frac{1}{2}(\frac{Log(u_h)}{\sigma})^2} \\ \\ \frac{1}{\sigma^2}[SM \ - 2SL\,Log(s) + (Log(s))^2] = \\ \\ 1 + \frac{1}{\sqrt{2\pi}(1-G(x_h))}\frac{Log(u_h)}{\sigma} e^{-\frac{1}{2}(\frac{Log(u_h)}{\sigma})^2} \end{cases}$$

Brunet-Moret l'écrit sous la forme :

$$Log(s) = \frac{(SM - SL\,Log(x_h) - \sigma^2)}{(SL - Log(x_h))} \tag{6}$$

$$SL = Log(s) + \sigma \frac{e^{-\frac{1}{2}(\frac{Log(u_h)}{\sigma})^2}}{\sqrt{2\pi}(1-G(x_h))}$$

et propose la méthode de résolution suivante. Elle consiste à choisir une valeur de σ, d'où une valeur de s d'après l'équation (6), puis à calculer :

$$SK = Log(s) + \sigma \frac{e^{-\frac{1}{2}\left(\frac{Log(u_h)}{\sigma}\right)^2}}{\sqrt{2\pi}(1 - G(x_h))}$$

en y portant les valeurs de σ et de s.

On compare ensuite SK et SL donné par la relation (5). Si SK est inférieur à SL, il faut faire décroître σ, sinon il faut faire croître σ. On continue les itérations jusqu'à avoir $SK/SL = 1$ avec l'erreur relative acceptée. $G(x_h)$ est obtenu à partir de la fonction de répartition de la loi normale réduite.

En effet

$$G(x_h) = \int_{-\infty}^{u_h} \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2}\left(\frac{Log(u)}{\sigma}\right)^2} du$$

soit :

$$G(x_h) = \int_{-\infty}^{\frac{Log(u_h)}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Références bibliographiques

Brunet-Moret (Y.), 1969. Etude de quelques lois statistiques utilisées en "Hydrologie". Cahier ORSTOM, série hydrologie, vol. 6, n° 3, 3-100.

Brunet-Moret (Y.), 1973. Estimation de paramètres. Cahier ORSTOM, série hydrologie, vol. 10, n° 2, 111-132.

Brunet-Moret (Y.), 1975. Distribution gausso-logarithmique. Cahier ORSTOM, série hydrologie, vol. 12, n° 2, 63-140.

Cohen (A.C.), 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics, vol. 1, n° 3, 217-237.

El-Shaarawi (A.H.), 1989. - Inferences about the mean from censored water quality data. Water Resources Research, vol. 25, n° 4, 685-690.

Gilliom (R.J.), Helsel (D.R.), 1986. Estimation of distributional parameters for censored trace level water quality data. 1. Estimation techniques. Water Resources Research, vol. 22, n° 2, 135-146.

Harter (H.L.), 1967. - Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population from complete and censored samples. Technometrics, vol. 9, n° 1, 159-165.

Hosking (J.R.M.), Wallis (J.R.), 1986. The value of historical data in flood frequency analysis. Water Resources Research, vol. 22, n° 11, 1606-1612.

Kendall (M.G.), Stuart (A.), 1977. - The advanced theory of statistics, vol. 1. Charles Griffin, Londres, 472p.

Kendal (M.G.), Stuart (A.), 1977. - The advanced theory of statistics, vol. 2. Charles Griffin, Londres, 690p.

Leese (M.N.), 1973. - Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. Water Resources Research, vol. 9, n° 6, 1534-1542.

Phien (H.N.), Fang (T.-S.E.), 1989. - Maximum-likelihood estimation of the parameters and quantiles of the general extreme-value distribution from censored samples. Journal of Hydrology, vol. 105, 139-155.