

Estimation de paramètres

Y. BRUNET-MORET

Ingénieur hydrologue à l'O.R.S.T.O.M.

RÉSUMÉ

Cet article concerne plus spécialement la détermination des paramètres d'ajustement à un échantillon d'une fonction de répartition continue et unimodale.

On commence par donner les définitions des estimateurs corrects et absolument corrects, efficaces et exhaustifs, avant d'exposer l'estimation par la méthode des moments et par la méthode du maximum de vraisemblance.

ABSTRACT

This paper particularly deals with estimation of fitting parameters of a continuous and unimodal distribution to a sample.

It begins by definitions of estimators consistent and unbiased, efficient and of minimum variance, then states estimation by moments' method and by maximum likelihood method.

SOMMAIRE

1. *Introduction*
2. *Méthodes d'estimation*
3. *Distribution limite d'un paramètre*
4. *Estimateurs corrects et absolument corrects*
 - 4.1. *Estimateurs corrects*
 - 4.2. *Estimateurs absolument corrects ou non biaisés*
 - 4.3. *Estimateurs biaisés*
 - 4.4. *Correction d'un estimateur biaisé*
 - 4.5. *Détermination du biais*
5. *Estimateurs efficaces et exhaustifs*
 - 5.1. *Cas d'un seul paramètre à déterminer*
 - 5.2. *Efficacité*
 - 5.3. *Cas de plusieurs paramètres à déterminer*
 - 5.4. *Retour sur l'efficacité*
6. *Méthode des moments*
 - 6.1. *Principe*
 - 6.2. *Variance des paramètres*
 - 6.3. *Variante par les cumulants*
 - 6.4. *Application : paramètres de position et d'échelle*
 - 6.5. *Formules relatives aux cumulants*
7. *Méthode du maximum de vraisemblance*
 - 7.1. *Principe*
 - 7.2. *Cas du (ou des) paramètres de position*
 - 7.3. *Cas d'un seul paramètre à estimer*
 - 7.4. *Cas de plusieurs paramètres à estimer*

- 7.5. Application : paramètres de position et d'échelle
- 7.6. Exemple : calcul des paramètres de position et d'échelle
- 8. Compléments
 - 8.1. Fonctions de paramètres
 - 8.2. Retour sur la méthode des moments
 - 8.3. Lois tronquées
 - 8.4. Troncatures
 - 8.5. Tronquage et troncature
- 9. Conclusion
 - 9.1. Remarques

1. INTRODUCTION

« Ce qui importe au statisticien, ce ne sont pas les valeurs observées mais ce qu'il en tire » : le statisticien cherche à condenser l'information diffuse dans la série d'observations qu'on lui soumet, et pour le faire, il est souvent amené à estimer des paramètres. On peut dire que l'ensemble des valeurs numériques de paramètres jouant un rôle bien déterminé, ainsi que les intervalles de confiance qui leur sont associés, résument l'information contenue dans un échantillon d'observations exprimées numériquement dans des unités spécifiées.

L'intervalle de confiance que peut fournir le statisticien n'a rien à voir avec la confiance que l'on peut avoir dans l'exactitude des observations et leur représentativité du réel. Dans un échantillon, il suffit d'un petit nombre de valeurs numériques soit disant observées mais trop fantaisistes pour conduire à des estimations complètement fausses de paramètres ; en effet, ces estimations sont alors entachées, d'une part de l'erreur due à l'échantillonnage (l'échantillon peut être mal représentatif de la population mère) et, d'autre part, des erreurs d'observations (dont on ne sait pas tenir compte en tant que telles sauf dans des cas très particuliers, et qui se confondent en pratique avec l'erreur d'échantillonnage).

Quand le statisticien écrit « l'intervalle de confiance à tant pour cent du module interannuel est compris entre telle et telle valeur », cela doit s'entendre : « étant donné le nombre de débits moyens annuels observés, la forme mathématique choisie pour représenter la fonction de probabilité de ces débits, le mode de calcul des paramètres ajustés de cette fonction », avec les restrictions : « sauf grossières erreurs sur les valeurs des débits » et « sauf mauvaise représentativité de l'échantillon qui peut avoir conduit à un mauvais choix de la loi de répartition de la population mère ».

2. MÉTHODES D'ESTIMATION

On se trouve en présence d'un échantillon de n observations, supposées indépendantes, et nous avons choisi la forme mathématique de la fonction de répartition devant représenter la loi de probabilité de la population mère de laquelle a été extrait l'échantillon. On a à calculer le mieux possible les valeurs numériques des paramètres en tenant compte, au mieux, de l'information contenue dans l'échantillon.

Les deux principales méthodes utilisées procèdent :

- par le maximum de vraisemblance,
- par les moments de l'échantillon.

Citons également, pour mémoire, la méthode de minimisation de la valeur d'un test d'adéquation : le seul test théoriquement utilisable est celui du χ^2 qui n'est pas assez consistant dans le cas des échantillons de petites tailles auxquels nous avons à faire pour qu'on puisse espérer s'en servir dans l'application de cette méthode de minimisation. Cependant, à la limite pour un échantillon de taille infiniment grande, cette méthode conduit à des estimations aussi correctes et efficaces que la méthode du maximum de vraisemblance.

La méthode des moindres carrés implique que les paramètres à déterminer soient des fonctions linéaires des observations (ou de transformations des observations n'incluant aucun des paramètres à déterminer). C'est le cas des régressions simples ou multiples, linéaires ou quelquefois non linéaires. Ce n'est jamais une méthode souhaitable pour estimer des paramètres de fonctions de répartition.

3. DISTRIBUTION LIMITE D'UN PARAMÈTRE

Soit un paramètre de valeur inconnue θ à déterminer, un échantillon de n valeurs indépendantes de la variate x et l'estimation du paramètre θ par

$$t_n = \varphi(x_1 \dots x_n)$$

La fonction φ des n valeurs observées provient soit de la définition du paramètre (moyenne, variance...), soit de la forme mathématique de la fonction de répartition choisie.

Si, à chaque échantillon aléatoire possible de n valeurs indépendantes de la variate, on associe la valeur correspondante de t_n , cette valeur peut être considérée comme une variable aléatoire dont la distribution est déterminée par la connaissance de la fonction φ .

Si l'espérance mathématique de t_n ne diffère pas de θ , ou diffère de θ d'un terme d'ordre $1/n$, et si sa variance est d'ordre $1/n$, l'inégalité de Bienaymé-Tchebycheff est d'application immédiate : lorsque n tend vers l'infini, la distribution de t_n tend vers une distribution normale de moyenne θ et de variance tendant vers zéro.

Cette proposition s'étend à la distribution conjointe de plusieurs paramètres : si chacun d'eux diffère de la valeur exacte d'un terme d'ordre $1/n$ et possède une variance d'ordre $1/n$, lorsque n tend vers l'infini, la distribution conjointe des paramètres tend vers une distribution multinormale de moyennes égales aux valeurs exactes et de variances tendant vers zéro. Comme les covariances sont, en général, d'ordre $1/n$ (de même que les variances) les coefficients de corrélation tendent vers des valeurs limites, et, asymptotiquement, les distributions des paramètres ne sont pas indépendantes.

4. ESTIMATEURS CORRECTS ET ABSOLUMENT CORRECTS

Un estimateur t_n calculé d'après un échantillon de taille n est consistant (ou convergent) si d'une part la variance de sa distribution d'échantillonnage décroît lorsque n croît, d'autre part si, θ étant dans la population mère la valeur exacte du paramètre à estimer, on peut déterminer une taille n d'échantillon telle que $|t_n - \theta| < \varepsilon$ avec une probabilité $1 - \eta$, ε et η choisis positifs et aussi petits que l'on veut.

4.1. ESTIMATEURS CORRECTS

On dit, lorsque l'inégalité ci-dessus est vérifiée, que t_n converge en probabilité vers θ : et l'estimateur est correct lorsque la valeur limite de l'espérance mathématique de t_n est θ lorsque n tend vers l'infini. L'estimateur correct peut donc être affecté d'un biais d'ordre $1/n$.

4.2. ESTIMATEURS ABSOLUMENT CORRECTS OU NON BIAISÉS

Un estimateur t_n calculé d'après un échantillon de taille n est absolument correct si, d'une part, la variance de sa distribution d'échantillonnage décroît lorsque n croît et si d'autre part, θ ayant la même signification que ci-dessus, l'espérance mathématique de t_n est θ quel que soit n .

L'estimateur absolument correct n'est pas biaisé.

4.3. ESTIMATEURS BIAISÉS

On peut, par exemple, trouver directement l'estimateur absolument correct de la variance μ_2 (moment centré d'ordre 2) lorsque la moyenne \bar{x} est calculée d'après l'échantillon de n valeurs x_i (m_2 étant le moment non centré d'ordre 2)

$$\begin{aligned} t_n &= \frac{E \left[\sum (x_j - \bar{x})^2 \right]}{n-1} = \frac{E \left[\sum \left(x_j - \frac{\sum x_j}{n} \right)^2 \right]}{n-1} = E \left[\frac{n-1}{n} \sum x_j^2 - \frac{1}{n} \sum_{j \neq k} \sum x_j x_k \right] / (n-1) \\ &= \frac{1}{n-1} E \left[\frac{n-1}{n} \sum x_j^2 - \sum_{k \neq j} x_j \frac{\sum x_k}{n} \right] = \frac{(n-1)m_2 - (n-1)\bar{x}^2}{n-1} = \mu_2 \end{aligned}$$

Un estimateur simplement correct de la variance, dans le cas d'une loi normale de paramètres de position et d'échelle à estimer peut être calculé par la méthode du maximum de vraisemblance :

$$t_n = \frac{1}{n} E \left[\sum (x_j - \bar{x})^2 \right] = \mu_2 \left(1 - \frac{1}{n} \right)$$

Cet estimateur t_n est correct et biaisé d'ordre $1/n$.

4.4. CORRECTION D'UN ESTIMATEUR BIAISÉ

θ étant la valeur exacte du paramètre dans la population mère, l'estimateur biaisé t_n , calculé d'après l'échantillon de taille n , peut s'écrire en fonction des cumulants déduits de l'échantillon (ces cumulants sont des estimateurs non biaisés des cumulants de la population mère supposés tous exister). Comme les moments des cumulants peuvent se développer en séries de puissance de $1/n$, on peut écrire :

$$E(t_n) - \theta = \sum_1^{\infty} \frac{a^i}{n^i}$$

t_n étant biaisé d'ordre $1/n$.

En appelant \bar{t}_{n-1} la valeur moyenne des n estimateurs t_{n-1} calculés (de la même façon que t_n) sur les n sous-échantillons différents possibles, comportant chacun $n-1$ valeurs extraites de l'échantillon original, et en posant

$$t'_n = n t_n - (n-1) \bar{t}_{n-1}$$

il vient :

$$E(t'_n) - \theta = a_2 \left(\frac{1}{n} - \frac{1}{n-1} \right) + a_3 \left(\frac{1}{n^2} - \frac{1}{(n-1)^2} + \dots \right)$$

et t'_n est biaisé d'ordre $1/n^2$

De même, en posant :

$$t''_n = [n^2 t'_n - (n-1)^2 \bar{t}'_{n-1}] / [n^2 - (n-1)^2]$$

on voit que t''_n est biaisé d'ordre $1/n^3$, etc.

Asymptotiquement, la variance de t'_n est la même que celle de t_n si celle-ci est de l'ordre de $1/n$, ce qui n'est pas vrai en général pour t''_n . Autrement dit, si t_n est exhaustif ou efficace (voir plus loin), t'_n sera efficace, t''_n , etc., risquent d'être moins efficaces que t'_n .

Il est peut-être possible d'utiliser la méthode ci-dessus, même si les cumulants de la population mère n'existent pas tous et lorsque les paramètres ne sont pas déterminés isolément, mais par un système d'équations implicites dont chacune en contient un ou plusieurs.

4.5. DÉTERMINATION DU BIAIS

Il semble que, dès qu'on estime plusieurs paramètres à la fois, quelle que soit la méthode d'estimation, on obtienne systématiquement des valeurs biaisées, probablement d'ordre $1/n$. Quand n est grand, cela n'a peut-être pas beaucoup d'importance, car les distributions sur lesquelles on travaille se modifient peu pour de petites variations des paramètres.

Lorsque n est petit, ces déformations peuvent devenir importantes.

Nous n'avons pas trouvé, dans la littérature statistique, de méthode permettant de mettre en évidence des biais. Il nous semble que seules des méthodes de Monte Carlo permettraient, en calculant t'_n , de déceler ces biais et de les corriger.

5. ESTIMATEURS EFFICACES ET EXHAUSTIFS

Il peut exister plus d'un estimateur absolument correct, non biaisé, du même paramètre.

Exemple : Dans le cas d'une loi normale de variance 1 :

$$dF(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2}} dx,$$

on peut estimer la valeur centrale x_0 par la moyenne des observations ou par la valeur médiane de l'échantillon (s'il est de taille impaire). Mais la variance de la moyenne est $1/n$ et celle de la médiane de l'ordre de $\frac{\pi}{2n}$; on dit que la moyenne est un estimateur plus « efficace » que la médiane.

Exemple : Dans le cas d'une loi de Cauchy de paramètre d'échelle 1 :

$$dF(x) = \frac{1}{\pi} \frac{dx}{1 + (x - x_0)^2}$$

la médiane, dont la variance est de l'ordre $\frac{\pi^2}{4n}$, est un meilleur estimateur de la valeur centrale que la moyenne dont la variance est infinie.

5.1. CAS D'UN SEUL PARAMÈTRE A DÉTERMINER

Soit θ le paramètre dont il faut estimer la valeur. Si les deux premières dérivées, par rapport à θ , de la fonction de vraisemblance (cf. 7.1.) existent, et si l'intervalle de définition de la variate est indépendant de la valeur de θ , on démontre que :

$$E \left[\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right)^2 \right] = - E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2} \right)$$

Si t est un estimateur sans biais d'une fonction $\tau(\theta)$ (t peut donc être un estimateur non biaisé de θ , ou un estimateur biaisé de θ par l'intermédiaire de la fonction $\tau(\theta)$), on montre que :

$$\text{var } t \geq \left[\frac{d\tau}{d\theta} \right]^2 / E \left[\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right)^2 \right]$$

la quantité à droite de l'inégalité ne tenant compte que des termes en $1/n$.

5.1.1. Un estimateur t dont la variance est égale à :

$$- \left[\frac{d\tau}{d\theta} \right]^2 / E \left[\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2} \right]$$

quel que soit n , est dit exhaustif : il possède la plus petite variance possible, étant données la taille de l'échantillon et la fonction de répartition.

Dans le cas où t est un estimateur de θ : $\frac{d\tau}{d\theta} = 1$ et la quantité : $I = E \left[\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right)^2 \right] = - E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2} \right)$ est appelée quantité d'information contenue dans l'échantillon.

5.1.2. L'estimateur t de la fonction $\tau(\theta)$ ne peut avoir la variance minimale (être exhaustif) que si l'on peut mettre $\frac{\partial \text{Log } \mathcal{L}}{\partial \theta}$ sous la forme $A(\theta)[t - \tau(\theta)]$, $A(\theta)$ étant une fonction de θ ne contenant aucune des valeurs des observations. Comme :

$$\text{var} \left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right) = E \left[\left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right)^2 \right] = [A(\theta)]^2 \text{var } t$$

il vient :

$$\text{var } t = \left| \frac{d\tau}{d\theta} / A(\theta) \right|$$

Exemple : Dans le cas d'une distribution normale de variance connue σ^2

$$dF(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} dx,$$

le paramètre de position θ se calcule par

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} = \frac{n}{\sigma^2} (\bar{x} - \theta) = 0$$

ce qui est la forme ci-dessus avec $t = \bar{x}$, $A(\theta) = \frac{n}{\sigma^2}$ et $\tau(\theta) = \theta$

\bar{x} est l'estimateur exhaustif de θ avec la variance $\frac{\sigma^2}{n}$

Exemple : Dans le cas d'une distribution de Cauchy de paramètre d'échelle connu 1,

$$dF(x) = \frac{1}{\pi} \frac{dx}{1 + (x - \theta)^2}$$

le paramètre de position se calcule par

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} = 2 \sum \frac{x_i - \theta}{1 + (x_i - \theta)^2} = 0$$

qui ne peut se mettre sous la forme $A(\theta)[t - \tau(\theta)]$: il n'y a pas d'estimateur exhaustif pour le paramètre de position.

5.1.3. Il s'ensuit que s'il existe un estimateur exhaustif pour une fonction $\tau(\theta)$ du paramètre θ , il n'en existe pas pour une autre fonction du même paramètre.

Exemple : Dans le cas d'une distribution normale à paramètre de position connu (moyenne), il existe un estimateur exhaustif de la variance, mais il n'existe pas d'estimateur exhaustif de la racine carrée de la variance, c'est-à-dire de l'écart-type (paramètre d'échelle).

5.1.4. On montre que les fonctions de densité pour lesquelles il est possible d'obtenir un estimateur exhaustif d'une fonction d'un paramètre θ doivent pouvoir se mettre sous la forme (dite de la famille exponentielle des distributions) :

$$f(x) = \exp [A(\theta) B(x) + C(x) + D(\theta)]$$

A et D étant fonctions de θ , ne contenant pas x

B et C étant fonctions de x , ne contenant pas θ

Exemple : Dans le cas d'une distribution de Gumbel de paramètre de position connu égal à zéro, la fonction de densité contenant l'unique paramètre d'échelle θ s'écrit

$$f(x) = \exp \left[\text{Log } \theta - \frac{x}{\theta} - e^{-x/\theta} \right]$$

et ne peut se mettre sous la forme de la famille exponentielle.

Exemple : Dans le cas de la même distribution de paramètre d'échelle connu égal à 1, la fonction de densité contenant l'unique paramètre de position θ s'écrit $f(x) = \exp [-x + \theta - e^{-x} e^{\theta}]$ ce qui est la forme de la famille exponentielle.

5.1.5. Dans le cas où il existe un estimateur exhaustif t d'une fonction $\tau(\theta)$ du paramètre θ , la variance de l'estimateur d'une autre fonction $\psi[\tau(\theta)]$ de θ tend vers

$$-\left(\frac{\partial \psi}{\partial \theta}\right)^2 / E\left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2}\right)$$

(les termes négligés étant d'ordre $1/n^2$ et plus) lorsque n tend vers l'infini. L'estimateur de cette nouvelle fonction du paramètre θ n'est pas exhaustif mais seulement efficace (voir plus loin).

5.1.6. L'inégalité

$$\text{var}(t) > -\left(\frac{d\tau}{d\theta}\right)^2 / E\left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2}\right)$$

est encore vraie lorsque les deux bornes, ou une seule borne, de l'intervalle de définition de la variate dépendent d'un seul paramètre θ , à condition que la fonction de densité et sa dérivée par rapport à θ soient nulles pour les valeurs de la variate correspondant aux bornes (ou à la seule borne) et pour la valeur de θ .

Cela impose à la fonction de densité d'avoir la forme « en cloche ». L'estimateur n'est pas exhaustif, il est seulement efficace.

5.2. EFFICACITÉ

Nous avons défini l'estimateur exhaustif qui, lorsqu'il existe, possède la plus petite variance possible, quelle que soit la taille de l'échantillon. Cet estimateur n'existe pas toujours, mais on peut souvent trouver un estimateur consistant dont la variance tend vers l'expression donnée en 5.1.1. pour l'estimateur exhaustif, lorsque la taille de l'échantillon tend vers l'infini : cet estimateur est dit efficace.

On mesure l'efficacité d'un autre estimateur consistant, par rapport à un estimateur efficace ou exhaustif, comme étant l'inverse du rapport des tailles de grands échantillons nécessaires pour obtenir des variances égales des deux estimateurs. Comme ces variances sont généralement d'ordre $1/n$, on calcule l'efficacité d'un estimateur consistant en divisant, par sa propre variance, la variance de l'estimateur efficace auquel on le compare.

Exemple : Dans le cas d'une distribution normale de variance connue σ^2 , la moyenne observée est un estimateur sans biais et exhaustif du paramètre de position, de variance $\frac{\sigma^2}{n}$. La médiane observée en est un estimateur consistant (et même absolument correct) de variance $\frac{\pi\sigma^2}{2n}$ (aux termes d'ordres supérieurs à $\left(\frac{1}{n}\right)^2$ près). L'efficacité de la médiane est $\frac{2}{\pi} = 0.637$.

5.3. CAS DE PLUSIEURS PARAMÈTRES A DÉTERMINER

Les résultats obtenus en 5.1. s'étendent au cas de plusieurs paramètres à déterminer simultanément.

$\theta_1, \theta_2, \dots, \theta_k$ étant les k paramètres à déterminer, t un estimateur sans biais d'une fonction $\tau(\theta_1, \dots, \theta_k)$ des paramètres, on a :

$$\text{var } t \geq \sum_{i=1}^k \sum_{j=1}^k \frac{\partial \tau}{\partial \theta_i} \frac{\partial \tau}{\partial \theta_j} I_{ij}^{-1}$$

à condition que les deux premières dérivées, par rapport aux paramètres, de la fonction de vraisemblance existent et que l'intervalle de définition de la variate soit indépendant des valeurs des paramètres θ .

Ceci peut se généraliser à un ensemble de $m \leq k$ estimateurs t_e non biaisés des m fonctions $\tau_e(\theta_1, \dots, \theta_k)$

$$\text{var } t_e \geq \sum_{i=1}^k \sum_{j=1}^k \frac{\partial \tau_e}{\partial \theta_i} \frac{\partial \tau_e}{\partial \theta_j} I_{ij}^{-1}$$

où I_{ij}^{-1} est le terme correspondant, dans la matrice inverse, du terme I_{ij} de la matrice d'information

$$I_{ij} = E \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta_i} \frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta_j} \right)$$

(nous retrouverons cette matrice dans la méthode du maximum de vraisemblance).

Ce qui précède peut encore se généraliser aux distributions multivariates.

5.3.1. Généralisation de 5.1.4.

La condition pour qu'une fonction de distribution puisse posséder un ensemble de k estimateurs exhaustifs de k fonctions des k paramètres à déterminer, est qu'elle puisse se mettre sous la forme :

$$dF(x) = \exp [D(\theta_1, \dots, \theta_k) + C(x) + \sum_1^k A_i(\theta_1, \dots, \theta_k) B_i(x)]$$

les k fonctions B et la fonction C ne contenant aucun des k paramètres θ ,

les k fonctions A et la fonction D ne contenant pas x , (ceci peut se généraliser aux distributions multivariates),

Exemple : la fonction de densité de la loi normale à deux paramètres de position x et d'échelle σ peut se mettre sous la forme de la famille exponentielle :

$$\exp \left[-\text{Log}(\sigma\sqrt{2\pi}) - \frac{\bar{x}^2}{2\sigma^2} + \frac{\bar{x}}{\sigma} x - \frac{x^2}{2\sigma^2} \right]$$

5.3.2. Il semble que l'on puisse généraliser et obtenir des estimateurs efficaces lorsque les deux bornes, ou une seule, de l'intervalle de définition de la variate dépendent de deux paramètres θ_h ou d'un seul, à condition que la fonction de densité et ses dérivées par rapport aux paramètres θ_h soient nulles pour les valeurs de la variate correspondant aux bornes (ou à la seule borne) et pour les valeurs des θ_h .

La fonction de densité a la forme « en cloche ».

5.4. RETOUR SUR L'EFFICACITÉ

Les définitions du paragraphe 5.2. sont encore valables dans le cas de plusieurs paramètres à déterminer, et la généralisation semble immédiate.

Mais il y a une difficulté de taille : quand on s'intéresse, ce qui semble assez normal, aux paramètres proprement dits de la distribution, et non à des fonctions de ces paramètres, fonctions qui sont inutilisables telles quelles. Les estimations des paramètres ne sont pas indépendantes, mais liées entre elles par des relations qui n'ont rien de linéaire.

Pour des tailles très grandes d'échantillons, on peut démontrer que les distributions conjointes des estimateurs des paramètres sont multinormales, et c'est dans ce cas seulement qu'on peut mesurer l'efficacité de telle méthode de détermination de tel paramètre par rapport à celle qui (en admettant qu'elle existe toujours) conduirait à la variance minimale : il n'est pas prouvé, dans le cas d'échantillons de petites tailles, que cette dernière méthode soit réellement la plus efficace (1). Il semble que seules des méthodes de Monte-Carlo permettraient de préciser cette question.

6. MÉTHODE DES MOMENTS

6.1. PRINCIPE

On calcule, à partir des valeurs observées, les moments « naturels » (non centrés) depuis l'ordre 1 (moyenne) jusqu'à l'ordre de rang égal au nombre de paramètres à estimer. Les moments ainsi calculés sont des estimateurs non biaisés des moments correspondants de la population mère, s'ils existent.

On détermine, d'après la fonction de répartition choisie, les équations donnant ces moments s'ils existent en fonction des paramètres. On obtient ainsi un système d'équations, souvent implicites et transcendantes, en nombre égal à celui des paramètres à estimer, système qu'il faut résoudre, par approximations successives si nécessaire.

Les estimations ainsi obtenues sont consistantes, mais non absolument correctes ni exhaustives. Leurs efficacités semblent d'autant meilleures que la loi de distribution de la variate se rapproche d'une loi normale.

6.2. VARIANCES DES PARAMÈTRES

Si un paramètre p peut s'écrire sous la forme d'une fonction de l moments

$$p = \varphi (m_1, \dots, m_l)$$

la fonction φ étant linéaire suivant les moments de chaque ordre, et les lois de distribution des moments de chaque ordre (déduites de la fonction de répartition de la population mère d'où est tiré l'échantillon, des valeurs des paramètres calculés d'après l'échantillon, et de la taille de l'échantillon) étant unimodales et de variances finies, la variance du paramètre p peut s'écrire approximativement, en s'en tenant aux termes d'ordre $1/n$:

$$\text{var} (p) = \sum \left(\frac{\partial \varphi}{\partial m_i} \right)^2 \text{var} (m_i) + 2 \sum_{k \neq i} \left(\frac{\partial \varphi}{\partial m_i} \right) \left(\frac{\partial \varphi}{\partial m_k} \right) \text{covar} (m_i, m_k)$$

avec :

$$\text{var} (m_i) = \frac{1}{n} (m_{2i} - m_i^2)$$

$$\text{covar} (m_i, m_k) = \frac{1}{n} (m_{i+k} - m_i m_k)$$

en prenant pour chacun de ces moments « m_j » la valeur du moment déduite de la fonction de répartition dans laquelle on utilise les valeurs calculées des paramètres.

Cette expression de la variance de la valeur du paramètre peut encore être utilisée lorsque la fonction φ n'est pas linéaire suivant les moments, lorsque l'échantillon est de grande taille.

(1) Sauf cas d'exhaustivité.

6.3. VARIANTE PAR LES CUMULANTS

On calcule à partir des valeurs observées les cumulants non biaisés de l'échantillon (formules ci-dessous), depuis l'ordre 1 (moyenne) jusqu'à l'ordre de rang égal au nombre de paramètres à estimer. On détermine, d'après la fonction de répartition choisie les équations donnant ces cumulants s'ils existent en fonction des paramètres : équations souvent plus simples que celles qui donnent les moments en fonction des paramètres : on obtient ainsi un système d'équations, souvent implicites et transcendantes, en nombre égal à celui des paramètres, système qu'il faut résoudre.

Les estimations ainsi obtenues sont consistantes mais ni absolument correctes ni exhaustives. Leur efficacité est d'autant meilleure que la loi de distribution se rapproche plus d'une loi normale.

Il est probable que les estimateurs calculés par les cumulants sont moins biaisés, bien que biaisés du même ordre en $1/n$, que les estimateurs calculés par les moments.

Il serait intéressant de vérifier cette assertion en utilisant des méthodes de Monte-Carlo.

Si un paramètre p peut s'écrire sous la forme d'une fonction de l cumulants

$$p = \varphi(k_1, \dots, k_l)$$

on peut calculer approximativement la variance de p (avec des restrictions analogues à celles qui ont été notées à propos du calcul par les moments) en s'en tenant aux termes d'ordre $1/n$,

$$\text{var}(p) = \sum \left(\frac{\partial \varphi}{\partial k_i} \right)^2 \text{var}(k_i) + 2 \sum_{j \neq i} \left(\frac{\partial \varphi}{\partial k_i} \right) \left(\frac{\partial \varphi}{\partial k_j} \right) \text{covar}(k_i, k_j)$$

les variances et covariances des cumulants étant données dans les formules ci-dessous (6.5.) dans lesquelles on doit utiliser les valeurs des cumulants déduites de la fonction de répartition de la population mère avec des valeurs calculées des paramètres. Les formules données ci-dessous pour les variances et covariances de cumulants n'ont pas le biais systématique des formules analogues données pour les moments (6.2.).

6.4. APPLICATION : PARAMÈTRES DE POSITION ET D'ÉCHELLE

Nous supposons n'avoir que ces deux paramètres à déterminer d'après un échantillon de taille n . Posons :

$$u = \frac{x - x_0}{s}$$

x_0 étant paramètre de position et s le paramètre d'échelle. La loi de probabilité ayant été choisie, nous connaissons les valeurs des moments, centrés ou non, et des cumulants de u .

6.4.1. Par la méthode des moments : en posant

$$S_1 = \frac{1}{n} \sum x_i \qquad S_2 = \frac{1}{n} \sum x_i^2$$

et en appelant m_1 la valeur moyenne de u , μ_2 , μ_3 et μ_4 les moments centrés de 2^e, 3^e et 4^e ordre de u dans la loi de probabilité choisie.

Le système d'équations donnant x_0 et s s'écrit :

$$\begin{aligned} s m_1 &= S_1 - x_0 \\ s^2 m^2 &= S_2 - 2 S_1 x_0 + x_0^2 \end{aligned}$$

$$\text{d'où :} \quad S = \frac{\sqrt{S_2 - S_1^2}}{\sqrt{\mu_2}} \qquad \text{et} \qquad x_0 = S_1 - \frac{m_1}{\sqrt{\mu_2}} \sqrt{S_2 - S_1^2}$$

les variances de ces déterminations sont, en ne conservant que les termes en $1/n$, pour des échantillons de grande taille :

$$\begin{aligned} \text{var}(s) &= \frac{s^2}{4n \mu_2^2} (\mu_4 - \mu_3^2) \\ \text{var}(x_0) &= \frac{s^2}{n} \left[\mu_2 - \frac{m_1}{\mu_2} \mu_3 + \frac{m_1^2}{4 \mu_2^2} (\mu_4 - \mu_3^2) \right] \end{aligned}$$

En partant de $S_1 = s m_1 + x_0$, nous pouvons écrire, pour de grands échantillons :

$$\text{var}(S_1) = m_1^2 \text{var}(s) + \text{var}(x_0) + 2 m_1 \text{covar}(x_0, s)$$

d'où :

$$\text{covar}(x_0, s) = \frac{s^2}{4n\mu_2} \left[2\mu_3 - \frac{m_1}{\mu_2} (\mu_4 - \mu_2^2) \right]$$

Si on fait un changement d'origine $v = u - u_0$, ce qui entraîne pour la moyenne de v : $m'_1 = m - u_0$, les moments centrés ne changent pas de valeur et $\text{covar}(x_0, s)$ s'annule (aux termes d'ordre $\left(\frac{1}{n}\right)^2$ et supérieurs près) pour :

$$m'_1 = \frac{2\mu_2 \mu_3}{\mu_4 - \mu_2^2}$$

on a alors :

$$\mu_0 = -\frac{2 \mu_2 \mu_3}{\mu_4 - \mu_2^2} + m_1$$

Dans le nouveau système de variable réduite «décagée», la variance du nouveau paramètre de position x'_0 (défini par $s m'_1 = S_1 - x'_0$) devient :

$$\text{var}(x'_0) = \frac{s^2}{n} \left[\mu_2 - \frac{\mu_3^2}{\mu_4 - \mu_2^2} \right]$$

Ceci impose, pour que $\text{var}(x'_0)$ soit positif, $\gamma_1^2 < \gamma_2 + 2$, avec

$$\gamma_2 = \frac{\mu_4 - 3\mu_3^2}{\mu_2^2}$$

coefficient d'aplatissement (toujours supérieur à -2) et $\gamma_1 = \mu_3/\mu_2^{3/2}$ coefficient d'asymétrie. Cette condition est toujours remplie quelle que soit la fonction de distribution.

Le coefficient de corrélation linéaire entre x'_0 et s est nul si on ne tient compte que des termes en $1/n$. Asymptotiquement, les distributions de x'_0 et de s sont normales et indépendantes.

Si la distribution de u est symétrique, μ_3 est nul, la transformation $v = u - m_1$ conduit à $\text{var}(x'_0) = \frac{s^2 \mu_2}{n}$ les conclusions ci-dessus n'étant pas modifiées.

6.4.2. Supposons que l'intervalle de définition de la variée soit borné inférieurement et/ou supérieurement (une borne étant d'habitude représentée par x_0 (ou $x'_0 + s u_0$) et l'autre, si elle existe, par $x_1 = x_0 + s$). Il se peut que des valeurs observées de la variée se trouvent en dehors de l'intervalle de définition calculé. On le constate souvent lorsque les valeurs des dérivées par rapport à u de la fonction de densité $f(u)$ pour les valeurs de u représentant les bornes, ne sont pas nulles pour un certain nombre d'ordres, au moins trois semble-t-il. Cela réduit beaucoup l'intérêt de la méthode des moments (ou des cumulants) pour le calcul de paramètres de position du type «bornes», lorsque la fonction de densité n'a pas franchement l'allure «en cloche», quel que soit le nombre de paramètres à calculer.

6.4.3. Variante par les cumulants : les deux premiers cumulants déduits de l'échantillon sont :

$$R_1 = \frac{1}{n} \sum x_i \quad \text{et} \quad R_2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]$$

En appelant $k_1 = m_1$, k_2 , k_3 et k_4 les quatre premiers cumulants de u dans la loi de probabilité choisie

$$s k_1 = R_1 - x_0 \quad s^2 k_2 = R_2$$

d'où :

$$s = \sqrt{\frac{R_2}{k_2}} \quad x_0 = R_1 - \frac{k_1}{\sqrt{k_2}} \sqrt{R_2}$$

les variances de ces déterminations sont, en ne conservant que les termes en $1/n$ pour des échantillons de grande taille :

$$\begin{aligned} \text{var}(s) &= \frac{s^2}{4n k_2^2} \left[k_4 + 2 \frac{n}{n-1} k_2^2 \right] \\ \text{var}(x_0) &= \frac{s^2}{n} \left[k_2 + \frac{k_1^2}{4k_2^2} \left(k_4 + \frac{2n}{n-1} k_2^2 \right) - \frac{k_3 k_1}{k_2} \right] \\ \text{covar}(x_0, s) &= \frac{s^2}{4n k_2} \left[2k_3 - \frac{k_1}{k_2} \left(k_4 + \frac{2n}{n-1} k_2^2 \right) \right] \end{aligned}$$

Nous pouvons faire, comme plus haut, un changement d'origine $u' = u + u_0$ qui permette d'annuler les termes en $1/n$ de covar (x_0, s) . La condition s'écrit :

$$\gamma_1^2 < \gamma_2 + \frac{2n}{n-1}$$

elle est toujours remplie, on retrouve les mêmes conclusions qu'au paragraphe 6.4.1.

6.5. FORMULES RELATIVES AUX CUMULANTS

L'échantillon étant de taille n , en posant :

$$S_j = \sum_{i=1}^n x_i^j$$

— Cumulants

$$k_1 = \frac{1}{n} S_1 \quad \text{moyenne}$$

$$k_2 = \frac{1}{n(n-1)} [nS_2 - S_1^2] \quad \text{variance}$$

$$k_3 = \frac{1}{n(n-1)(n-2)} [n^2 S_3 - 3nS_2 S_1 + 2 S_1^3]$$

$$k_4 = \frac{1}{n(n-1)(n-2)(n-3)} [(n^3 + n^2)S_4 - 4(n^2 + n)S_3 S_1 - 3(n^2 - n)S_2^2 + 12nS_2 S_1^2 - 6S_1^4]$$

$$k_5 = \frac{1}{n(n-1)(n-2)(n-3)(n-4)} [(n^4 + 5n^3)S_5 - 5(n^3 + 5n^2)S_4 S_1 - 10(n^3 - n^2)S_3 S_2 + 20(n^2 + 2n)S_3 S_1^2 + 30(n^2 - n)S_2^2 S_1 - 60nS_2 S_1^3 + 24S_1^5]$$

— Variances :

$$\text{var}(k_1) = \frac{k_2}{n}$$

$$\text{var}(k_2) = \frac{k_4}{n} + \frac{2k_2^2}{n-1}$$

$$\text{var}(k_3) = \frac{k_6}{n} + \frac{9k_4 k_2}{n-1} + \frac{9k_3^2}{n-1} + \frac{6n}{n-1} \frac{k_2^3}{n-2}$$

$$\text{var}(k_4) = \frac{k_8}{n} + \frac{16 k_6 k_2}{n-1} + \frac{48 k_5 k_3}{n-1} + \frac{34 k_4^2}{n-1} + \frac{72n}{n-1} \frac{k_4 k_2^2}{n-2} + \frac{144n}{n-1} \frac{k_3^2 k_2}{n-2} + \frac{24n}{n-1} \frac{n+1}{n-2} \frac{k_2^4}{n-3}$$

$$\begin{aligned} \text{var}(k_5) &= \frac{k_{10}}{n} + \frac{25 k_8 k_2}{n-1} + \frac{100 k_7 k_3}{n-1} + \frac{200 k_6 k_4}{n-1} + \frac{125 k_5^2}{n-1} + \frac{200n}{n-1} \frac{k_6 k_2^2}{n-2} + \frac{1200n}{n-1} \frac{k_5 k_3 k_2}{n-2} + \frac{850n}{n-1} \frac{k_4^2 k_2}{n-2} \\ &+ \frac{1500n}{n-1} \frac{k_4 k_3^2}{n-2} + \frac{600n}{n-1} \frac{n+1}{n-2} \frac{k_4 k_2^3}{n-3} + \frac{1800n}{n-1} \frac{n+1}{n-2} \frac{k_3^2 k_2^2}{n-3} + \frac{120n^2}{n-1} \frac{n+5}{n-2} \frac{k_2^5}{(n-3)(n-4)} \end{aligned}$$

— Covariances :

$$\text{covar}(k_2, k_1) = \frac{k_3}{n}$$

$$\text{covar} (k_3, k_1) = \frac{k_4}{n}$$

$$\text{covar} (k_4, k_1) = \frac{k_5}{n}$$

$$\text{covar} (k_5, k_1) = \frac{k_6}{n}$$

$$\text{covar} (k_3, k_2) = \frac{k_5}{n} + \frac{6 k_3 k_2}{n-1}$$

$$\text{covar} (k_4, k_2) = \frac{k_6}{n} + \frac{8 k_4 k_2}{n-1} + \frac{6 k_3^2}{n-1}$$

$$\text{covar} (k_5, k_2) = \frac{k_7}{n} + \frac{10 k_5 k_2}{n-1} + \frac{20 k_4 k_3}{n-1}$$

$$\text{covar} (k_4, k_3) = \frac{k_7}{n} + \frac{12 k_6 k_2}{n-1} + \frac{30 k_4 k_3}{n-1} + \frac{36n}{n-1} \frac{k_3 k_2^2}{n-2}$$

$$\text{covar} (k_5, k_3) = \frac{k_8}{n} + \frac{15 k_6 k_2}{n-1} + \frac{45 k_5 k_3}{n-1} + \frac{30 k_4^2}{n-1} + \frac{60n}{n-1} \frac{k_4 k_3^2}{n-2} + \frac{90n}{n-1} \frac{k_3^2 k_2}{n-2}$$

$$\begin{aligned} \text{covar} (k_5, k_4) = & \frac{k_9}{n} + \frac{20 k_7 k_2}{n-1} + \frac{70 k_6 k_3}{n-1} + \frac{120 k_5 k_4}{n-1} + \frac{120n}{n-1} \frac{k_5 k_2^2}{n-2} + \frac{600n}{n-1} \frac{k_4 k_3 k_2}{n-2} + \frac{180n}{n-1} \frac{k_3^3}{n-2} \\ & + \frac{240n}{n-1} \frac{n+1}{n-2} \frac{k_3 k_3^2}{n-3} \end{aligned}$$

7. MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

7.1. PRINCIPE

Supposons qu'un échantillon tiré d'une population mère comporte n valeurs x_i pouvant se produire chacune avec la probabilité P_i . La probabilité pour qu'un échantillon de n valeurs obtenues par tirages indépendants soit précisément l'échantillon observé est :

$$P_1 P_2 \dots P_i \dots P_n$$

On appelle cette probabilité «vraisemblance de l'échantillon».

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

Si la variate est continue, chacun des termes ci-dessus, et le produit lui-même sont infiniment petits. On définit alors la vraisemblance de l'échantillon comme une quantité proportionnelle au produit des densités de probabilités, c'est-à-dire à :

$$\mathcal{L} = f_1 f_2 \dots f_i \dots f_n$$

(fonction de vraisemblance)

avec :

$$f_i = f(x_i, \theta_1, \dots, \theta_k)$$

$\theta_1, \dots, \theta_k$ étant les paramètres de la fonction de densité f , indépendants les uns des autres, dont les valeurs sont à déterminer.

La méthode consiste à maximiser \mathcal{L} , donc à annuler les dérivées partielles de \mathcal{L} par rapport aux différents paramètres de valeur inconnue, ce qui donne un système de k équations :

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = 0 \quad \dots \quad \frac{\partial \mathcal{L}}{\partial \theta_k} = 0$$

Il est plus simple d'écrire

$$\text{Log } \mathcal{L} = \sum_1^n \text{Log } f_i$$

et de remplacer le système ci-dessus par :

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \theta_j} = \frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta_j}$$

c'est-à-dire :

$$\sum_{i=1}^n \frac{1}{f_i} \frac{\partial f_i}{\partial \theta_j} = 0 \quad \text{pour } j = 1 \text{ à } k$$

7.2. CAS DU (OU DES) PARAMÈTRE DE POSITION

En écrivant

$$u_i = \frac{x_i - x_0}{s},$$

x_0 étant le paramètre de position et s le paramètre d'échelle, la dérivée de $\text{Log } \mathcal{L}$ par rapport à x_0 s'écrit :

$$\frac{\partial \text{Log } \mathcal{L}}{\partial x_0} = -\frac{1}{s} \sum \frac{f'(u_i)}{f(u_i)}$$

$f(u)$ est toujours positif non nul dans l'intervalle de définition de u , du moins dans les cas de fonctions de densité qui nous intéressent.

Pour que l'équation

$$\frac{\partial \text{Log } \mathcal{L}}{\partial x_0} = 0$$

ait un sens, il faut que $f'(u)$ existe pour toutes les valeurs de u (y compris pour les valeurs des bornes de l'intervalle de définition de u). Pour qu'il y ait un maximum de vraisemblance, il faut que

$$\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0^2}$$

existe et soit négatif pour la valeur de x_0 qui annule

$$\frac{\partial \text{Log } \mathcal{L}}{\partial x_0}$$

Il faut donc que $f'(u)$ puisse changer de signe dans l'intervalle de définition de u : c'est-à-dire, pour les fonctions de densité $f(u)$ continues, qu'il y ait un mode réel (valeur de u pour laquelle la dérivée de $f(u)$ par rapport à u est nulle, et la dérivée seconde de $f(u)$ par rapport à u est négative), et ceci pour une valeur de u qui soit différente de celles des bornes de son intervalle de définition.

7.3. CAS D'UN SEUL PARAMÈTRE A ESTIMER

7.3.1. S'il y a un seul paramètre à estimer, si l'intervalle de définition de la variate ne dépend pas de ce paramètre, si l'on peut dériver deux fois la fonction de vraisemblance \mathcal{L} par rapport à ce paramètre

— et si l'estimateur du paramètre θ , solution de l'équation

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} = 0$$

n'est pas biaisé, il est exhaustif et sa variance est :

$$\text{var } \theta = -1/E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta^2} \right) = -1/ \left[nE \left(\frac{\partial^2 \text{Log } f}{\partial \theta^2} \right) \right]$$

— si l'estimateur est biaisé, sa variance a , asymptotiquement, l'expression ci-dessus.

7.3.2. Si le seul paramètre à estimer définit une borne (ou les deux bornes) de l'intervalle de définition de la variate, et s'il peut se calculer par la méthode du maximum de vraisemblance, son estimation par une autre méthode peut lui associer une variance inférieure à celle qui provient de la méthode du maximum de vraisemblance. Mais

si on peut dériver une seconde fois la fonction de vraisemblance par rapport au paramètre, et si cette dérivée seconde a encore un sens pour la valeur de u correspondant à la borne (ou les valeurs de u correspondant aux bornes), on montre que la variance asymptotique de l'estimateur a la même expression que plus haut ; l'estimateur est alors efficace.

7.3.3. Lorsque l'équation du maximum de vraisemblance ne donne pas l'estimateur du paramètre sous une forme explicite, on peut, pour le calculer, utiliser le système suivant d'approximations successives, θ_{i+1} étant la $(i+1)$ ième approximation

$$\theta_{i+1} = \theta_i + \left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} \right)_{\theta_i} / (\text{var } \theta)_{\theta_i}$$

dans beaucoup de cas $\text{var } \theta$ pourra être remplacée par sa valeur asymptotique

$$\text{var } \theta = -1 / \left[n \text{E} \left(\frac{\partial^2 \text{Log } f}{\partial \theta^2} \right) \right]$$

(lorsque la taille de l'échantillon n'est pas grande, ce système peut ne pas converger).

Exemple : Dans le cas d'une distribution de Cauchy de paramètre d'échelle connu et égal à 1

$$dF(x) = \frac{1}{\pi} \frac{dx}{1 + (x - \theta)^2},$$

le paramètre de position θ est solution de

$$\frac{\partial \text{Log } \mathcal{L}}{\partial \theta} = 2 \sum \frac{x_i - \theta}{1 + (x_i - \theta)^2} = 0$$

équation de degré $(2n - 1)$ en θ . On calcule aisément :

$$-\frac{1}{\text{var } \theta} = n \text{E} \left(\frac{\partial^2 \text{Log } f}{\partial \theta^2} \right) = \frac{n}{\pi} \int_{-\infty}^{+\infty} \frac{2(x - \theta)^2 - 2}{[1 + (x - \theta)^2]^3} dx = -\frac{n}{2}$$

Lorsque l'échantillon est de taille assez grande, une première approximation de θ est la médiane observée.

7.4. CAS DE PLUSIEURS PARAMÈTRES A ESTIMER

7.4.1. Si l'intervalle de définition de la variée ne dépend d'aucun des paramètres à estimer, si l'on peut dériver deux fois la fonction de vraisemblance \mathcal{L} par rapport à ces paramètres, les estimateurs des paramètres, solutions du système d'équations

$$\sum_1^n \frac{1}{f_i} \frac{\partial f_i}{\partial \theta_j} = 0$$

sont en général biaisés mais corrects et efficaces. On ne peut obtenir que leurs variances asymptotiques, et les estimations ne sont pas indépendantes les unes des autres : la distribution conjointe des estimateurs tend asymptotiquement vers une distribution multinormale, dont la matrice des variances et covariances est l'inverse de la matrice d'information dont le terme I_{ij} est :

$$I_{ij} = \text{E} \left(\frac{\partial \text{Log } \mathcal{L}}{\partial \theta_i} \frac{\partial \text{Log } \mathcal{L}}{\partial \theta_j} \right) = - \text{E} \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta_i \partial \theta_j} \right)$$

7.4.2. Considérons un échantillon de taille fini. Si l'intervalle de définition de la variée dépend d'un (ou de deux) paramètre θ_b , si l'on peut dériver deux fois la fonction de vraisemblance \mathcal{L} par rapport à tous les paramètres (entre autres si les $\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta_b^2}$ et les $\frac{\partial^2 \text{Log } \mathcal{L}}{\partial \theta_i \partial \theta_b}$ existent) les conclusions ci-dessus restent valables ; cependant, l'estimation des paramètres θ_b par une autre méthode peut leur assurer des variances inférieures à celles auxquelles conduit le maximum de vraisemblance.

7.4.3. Le système d'équations du maximum de vraisemblance conduit rarement à des estimateurs explicites des paramètres, et il n'y a pas de méthode générale pour le résoudre.

Notons que, si l'intervalle de définition de la variate est borné, le mode de calcul du (ou des) paramètre représentant la (ou les deux) borne ne permet pas (même dans le cas où les dérivées secondes de $\log \mathcal{L}$ par rapport à ces paramètres n'existent pas) une estimation de la (ou des deux) borne à l'intérieur du domaine observé de variation de la variate. Le calcul de la borne inférieure conduit à une estimation inférieure ou égale à la plus petite des valeurs observées, le calcul de la borne supérieure conduit à une estimation supérieure ou égale à la plus grande des valeurs observées.

7.5. APPLICATION : PARAMÈTRES DE POSITION ET D'ÉCHELLE

7.5.1. Soit une fonction de distribution définie par

$$dF(x) = f\left(\frac{x-x_0}{s}\right) d\left(\frac{x-x_0}{s}\right),$$

on pose $u = \frac{x-x_0}{s}$ où x_0 est le paramètre de position et s le paramètre d'échelle, [$s = x_1 - x_0$ si x_1 est un second paramètre de position]. Le système d'équations du maximum de vraisemblance s'écrit en posant $g(u) = \text{Log } f(x)$

$$\begin{aligned}\frac{\partial \text{Log } \mathcal{L}}{\partial x_0} &= -\frac{1}{s} \sum g'(u_i) = 0 \\ \frac{\partial \text{Log } \mathcal{L}}{\partial s} &= -\frac{1}{s} \sum u_i g'(u_i) - \frac{n}{s} = 0\end{aligned}$$

en supposant que $g'(u)$ et $ug'(u)$ existent pour toutes les valeurs de u comprises dans son intervalle de définition. On démontre que

$$\begin{aligned}\mathbb{E}\left(\frac{\partial \text{Log } \mathcal{L}}{\partial x_0}\right) &= -\frac{n}{s} \mathbb{E}[g'(u)] = 0 \\ \mathbb{E}\left(\frac{\partial \text{Log } \mathcal{L}}{\partial s}\right) &= -\frac{n}{s} \mathbb{E}[ug'(u) + 1] = 0\end{aligned}$$

pour toutes les valeurs possibles de x_0 et de s (il faut que $f(u)$ soit nulle aux bornes), ce qui donne

$$\mathbb{E}[g'(u)] = 0 \qquad \mathbb{E}[ug'(u)] = -1$$

Les éléments de la matrice d'information sont :

$$\begin{aligned}- \mathbb{E}\left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0^2}\right) &= -\frac{n}{s^2} \mathbb{E}[g''(u)] \quad \text{à condition que } g''(u) \text{ existe aux bornes de } u; \\ - \mathbb{E}\left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0 \partial s}\right) &= \frac{n}{s^2} \mathbb{E}[g'(u) + ug''(u)] = -\frac{n}{s^2} \mathbb{E}[ug''(u)] \\ - \mathbb{E}\left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial s^2}\right) &= -\frac{n}{s^2} \mathbb{E}[u^2 g''(u) + 2ug'(u) + 1] = -\frac{n}{s^2} \mathbb{E}[u^2 g''(u) - 1]\end{aligned}$$

et la matrice d'information s'écrit :

$$-\frac{n}{s^2} \mathbb{E} \begin{bmatrix} g''(u) & ug''(u) \\ ug''(u) & u^2 g''(u) - 1 \end{bmatrix}$$

la matrice des variances et covariances devient :

$$v = -\frac{s^2}{n\Delta} \mathbb{E} \begin{bmatrix} u^2 g''(u) - 1 & -ug''(u) \\ -ug''(u) & g''(u) \end{bmatrix}$$

avec :

$$\Delta = E [g''(u)] E [u^2 g''(u) - 1] - (E [u g''(u)])^2$$

La variance (limite) de x_0 est :

$$- \frac{s^2}{n\Delta} E [u^2 g''(u) - 1]$$

La variance (limite) de s est :

$$- \frac{s^2}{n\Delta} E [g''(u)]$$

La covariance (limite) de x_0 et de s est $\frac{s^2}{n\Delta} E [u g''(u)]$.

7.5.2. Si la distribution est symétrique, $E[ug''(u)] = 0$, la covariance est de l'ordre de $\frac{1}{n^2}$, c'est-à-dire que les paramètres x_0 et s sont asymptotiquement indépendants.

Même si la distribution n'est pas symétrique, nous pouvons, par un changement d'origine, rendre nulle la covariance-limite de x_0 et s (comme dans la méthode des moments ou cumulants); posons :

$$v = u - \frac{E [u g''(u)]}{E [g''(u)]} = u - u_0$$

on peut écrire :

$$E[ug''(u)] = E[(v + u_0)g''(u)] = E[v g''(v)] + \frac{E[u g''(u)]}{E[g''(u)]} E[g''(u)] \quad \text{d'où} \quad E[v g''(v)] = 0$$

les éléments de la matrice d'information deviennent :

$$\begin{aligned} - E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0^2} \right) &= - \frac{n}{s^2} E [g''(v)] = - \frac{n}{s^2} E [g''(u)] \\ - E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial s^2} \right) &= \frac{n}{s^2} E [v^2 g''(v) - 1] = - \frac{n}{s^2} \left[E[u^2 g''(u) - 1] - \frac{[E [u g''(u)]]^2}{E[g''(u)]} \right] \end{aligned}$$

les variances (limites) de x_0 et de s sont les inverses des éléments ci-dessus.

7.6. *Exemple* : Calcul des paramètres de position et d'échelle d'après un échantillon de taille n dans le cas d'une loi gamma incomplète dont le paramètre de forme γ est connu a priori et supérieur à 2 :

$$dF(x) = \frac{1}{\Gamma(\gamma)} \left(\frac{x - x_0}{s} \right)^{\gamma-1} e^{-\frac{x - x_0}{s}} d \left(\frac{x - x_0}{s} \right) \quad \text{où} \quad s > 0$$

en posant $u = \frac{x - x_0}{s}$, on a :

$$\text{Log } f(u) = - \text{Log } \Gamma(\gamma) + (\gamma - 1) \text{Log } u - u = g(u)$$

7.6.1. Calcul des paramètres pour le maximum de vraisemblance :

$$\text{Log } \mathcal{L} = - n \text{Log } \Gamma(\gamma) - n\gamma \text{Log } s + (\gamma - 1) \sum \text{Log } (x_i - x_0) - \frac{1}{s} \sum (x_i - x_0)$$

dérivée par rapport à x_0 (condition $\gamma > 1$) :

$$\frac{\partial \text{Log } \mathcal{L}}{\partial x_0} = - (\gamma - 1) \sum \frac{1}{x_i - x_0} + \frac{n}{s} = 0$$

dérivée par rapport à s :

$$\frac{\partial \text{Log } \mathcal{L}}{\partial s} = - \frac{\gamma n}{s} + \frac{1}{s^2} \sum (x_i - x_0) = 0$$

d'où :

$$s = \frac{\Sigma(x_i - x_0)}{\gamma} = \frac{\bar{x} - x_0}{\gamma}$$

et :

$$\frac{\gamma - 1}{n} \sum \frac{1}{x_i - x_0} = \frac{\gamma}{\bar{x} - x_0}$$

équation de degré n en x_0 . Cette équation peut avoir n racines réelles : si les x_i sont rangés en ordre croissant et si nous faisons varier x_0 de x_1 à x_{i+1} , le premier membre varie de $-\infty$ à $+\infty$; si \bar{x} n'est pas compris entre x_i et x_{i+1} le second membre garde une valeur finie : il y a donc une racine entre x_i et x_{i+1} et $n - 2$ racines réelles, entre x_2 et x_n . En effet, si \bar{x} est compris entre x_i et x_{i+1} , il n'y a pas de racine dans cet intervalle. Il existe d'autre part une racine x_0 supérieure à x_n , plus grande valeur des x_i , et une racine inférieure à x_1 , plus petite valeur des x_i . Cette dernière est la seule à conserver parce que c'est la seule pour laquelle $\text{Log } \mathcal{L}$ ait une signification. En effet, $\text{Log}(x_i - x_0)$ n'a de sens que si x_0 est toujours inférieur à x_i .

Nous allons calculer les termes de la matrice d'information. Comme

$$g'(u) = \frac{\gamma - 1}{u} - 1 \quad \text{et} \quad g''(u) = -\frac{\gamma - 1}{u^2}$$

(condition $\gamma > 2$)

$$- E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0^2} \right) = -\frac{n}{s^2} E [g''(u)] = -\frac{n}{s^2} \int_0^\infty g''(u) f(u) du = \frac{n}{s^2} \frac{1}{\gamma - 2}$$

$$- E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0 \partial s} \right) = -\frac{n}{s^2} E [u g''(u)] = -\frac{n}{s^2} \int_0^\infty u g''(u) f(u) du = \frac{n}{s^2}$$

$$- E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial s^2} \right) = \frac{n}{s^2} [E[u^2 g''(u)] - 1] = -\frac{n}{s^2} \int_0^\infty u^2 g''(u) f(u) du + \frac{n}{s^2} = \frac{n}{s^2} \gamma$$

le déterminant de la matrice d'information a pour valeur :

$$\Delta = \frac{n^2}{s^4} \left[\frac{\gamma}{\gamma - 2} - 1 \right] = \frac{n^2}{s^4} \frac{2}{\gamma - 2}$$

d'où la matrice des variances et des covariances :

$$\frac{\gamma - 2}{2} \frac{s^2}{n} \begin{bmatrix} \gamma & -1 \\ -1 & \frac{1}{\gamma - 2} \end{bmatrix}$$

et : $\text{var}(x_0) = \frac{s^2}{n} \frac{\gamma(\gamma - 2)}{2}$ (variance limite)

$\text{var}(s) = \frac{s^2}{2n}$ (variance limite)

$\text{covar}(x_0, s) = \frac{s^2}{2n} (\gamma - 2)$ (covariance limite)

Le coefficient de corrélation linéaire entre x_0 et s tend asymptotiquement vers :

$$-\sqrt{\frac{\gamma - 2}{\gamma}}$$

En faisant le changement d'origine

$$v = u - \frac{E[ug''(u)]}{E[g''(u)]} = u - (\gamma - 2)$$

on obtient :

$$\text{var}(x'_0) = \frac{s^2}{n} (\gamma - 2) \quad \text{et} \quad \text{var}(s) = \frac{s^2}{2n} \quad (\text{variances limites})$$

avec :

$$\text{covar}(x'_0, s) = 0 \quad (\text{covariance limite})$$

Remarquons que si γ et s sont connus à priori

$$\text{var}(x_0) = -1/E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial x_0^2} \right) = \frac{s^2}{n} (\gamma - 2) \quad (\text{variance limite})$$

ou que si γ et x_0 sont connus à priori :

$$\text{var}(s) = -1/E \left(\frac{\partial^2 \text{Log } \mathcal{L}}{\partial s^2} \right) = \frac{s^2}{\gamma n} \quad (\text{variance limite})$$

7.6.2. Détermination par la méthode des moments

$$\begin{aligned} x_0 + \gamma s &= m_1 & s &= \sqrt{\frac{m_2 - m_1^2}{\gamma}} \\ \gamma s^2 &= m_2 - m_1^2 & x_0 &= m_1 - \sqrt{\gamma(m_2 - m_1^2)} \end{aligned}$$

La variance de s est :

$$\text{var}(s) = \frac{s^2}{2n} \left(1 + \frac{3}{\gamma} \right) \quad (\text{variance limite})$$

celle de x_0 est :

$$\text{var}(x_0) = \frac{s^2}{2n} \gamma(\gamma + 1) \quad (\text{variance limite})$$

et la covariance est :

$$\text{covar}(x_0, s) = -\frac{s^2}{2n} (\gamma + 1) \quad (\text{covariance limite})$$

le coefficient de corrélation linéaire entre x_0 et s est asymptotiquement de :

$$-\sqrt{\frac{\gamma + 1}{\gamma + 3}}$$

L'efficacité de la méthode des moments se trouve être :

— pour x_0 égale à : $\frac{\gamma - 2}{\gamma + 1}$ (pour de grands échantillons)

— pour s égale à : $\frac{\gamma}{\gamma + 3}$ (pour de grands échantillons)

En faisant le changement d'origine $v = u - \frac{\gamma(\gamma + 1)}{\gamma + 3}$ on obtient :

$$\text{var}(x'_0) = \frac{s^2}{n} \frac{\gamma(\gamma + 1)}{\gamma + 3} \quad \text{var}(s) = \frac{s^2}{2n} \frac{\gamma + 3}{\gamma} \quad (\text{variance limite})$$

avec $\text{covar}(x'_0, s) = 0$ (covariance limite).

8. COMPLÉMENTS

8.1. FONCTIONS DE PARAMÈTRES

Si Φ est une fonction des paramètres $\theta_1 \dots \theta_k$ l'estimation de Φ par le maximum de vraisemblance est obtenue en remplaçant dans la relation fonctionnelle les paramètres par leurs estimateurs obtenus par la méthode du maximum de vraisemblance, ce qui donne la variance asymptotiquement minimale à l'estimation de Φ .

8.2. RETOUR SUR LA MÉTHODE DES MOMENTS

Les estimations par la méthode simple des moments sont nettement plus biaisés (au moins dans le cas cité) que les estimations par la méthode du maximum de vraisemblance.

Dans certains cas, relativement simples, il est possible d'améliorer les estimateurs obtenus par la méthode des moments en tenant compte des covariances entre moments. L'expérience montre que les estimateurs ainsi améliorés des paramètres de forme et d'échelle d'une distribution gamma incomplète de paramètre de position connu semblent perdre beaucoup de leur biais et être aussi corrects que les estimateurs obtenus par la méthode du maximum de vraisemblance.

8.3. LOIS TRONQUÉES

8.3.1. Cas général : soit une fonction de répartition définie par $dG(x) = g(x)dx$ l'intervalle de définition étant $\alpha \leq x \leq \beta$, avec $\int_{\alpha}^{\beta} g(x)dx = 1$ (α et β pouvant être infinis en valeur absolue). On peut définir une fonction de répartition $F(x)$ telle que :

$$\begin{aligned}
 F(x) &= F_0 & \text{si} & \quad x < a & \quad a \geq \alpha \\
 F(x) &= F_1 & \text{si} & \quad x > b & \quad b \leq \beta \\
 F(x) &= F_0 + \frac{F_1 - F_0}{G} \int_a^x g(x)dx & \text{si} & \quad a \leq x \leq b & \quad \text{avec} \quad G = \int_a^b g(x)dx
 \end{aligned}$$

$f(x) = \frac{F_1 - F_0}{G} g(x)$ lorsque x se trouve dans l'intervalle (a, b) .

On peut dire que la loi est tronquée «en fréquence» et «en intervalle de définition».

Il peut y avoir k paramètres inclus dans $g(x) : \theta_1 \dots \theta_k$, les bornes a et b peuvent être des paramètres (ou des fonctions de paramètres), de même que les fréquences F_0 et F_1 .

L'application de la méthode des moments n'est simple que si on sait calculer

$$\int_a^b x^i g(x)dx = m_i(a, b, \theta_1, \dots, \theta_k) \quad \text{depuis} \quad i = 0$$

sans connaître à priori les valeurs des paramètres $a, b, \theta_1 \dots \theta_k$. Le système d'équations s'écrit, la taille n de l'échantillon étant connue et n_j étant le nombre d'observations x_j comprises entre a et b :

$$\frac{1}{n} \sum_{x_j \geq a}^{x_j \leq b} x_j^i = (F_1 - F_0) \frac{m_i(a, b, \theta_1, \dots, \theta_k)}{G(a, b, \theta_1, \dots, \theta_k)}$$

on ne peut donc pas déterminer séparément F_0 et F_1 par les moments mais seulement $(F_1 - F_0) = n_j/n$. F_0 et F_1 peuvent se déterminer d'après l'échantillon en prenant :

$$F_0 = \frac{\text{nombre d'observations} < a}{n}$$

ou

$$F_1 = \frac{\text{nombre d'observations} > b}{n}$$

L'application de la méthode du maximum de vraisemblance conduit à écrire :

$$\text{Log } \mathcal{L} = -n_j \text{Log} \left[\int_a^b g(x)dx \right] + \sum_{x_j \geq a}^{x_j \leq b} \text{Log} [g(x_j)]$$

sauf dans des cas très particuliers, le système d'équations qui s'en déduira par dérivation sera difficile à résoudre. Si F_0 et F_1 sont des paramètres indépendants de $a, b, \theta_1, \dots, \theta_k$, ils peuvent se déterminer d'après l'échantillon par les moments.

8.3.2. Le seul cas particulier de lois tronquées utilisées en hydrologie se réduit à un schéma très simple : il n'y a pas tronquage de l'intervalle de définition.

$$a = \alpha \quad \text{et} \quad b = \beta$$

De plus, ces valeurs sont connues à priori. La loi est tronquée en fréquence seulement, d'un seul côté, $F_1 = 1$, donc $G = 1$ et $f(x) = (1 - F_0)g(x)$.

Les applications des méthodes de détermination des paramètres (qui se réduisent aux paramètres d'échelle et de forme de $g(x)$ et au paramètre F_0) sont simples.

8.4. TRONCATURES

8.4.1. Il y a troncature de premier type des observations (en valeur) lorsqu'on ne peut observer les valeurs de la variate que dans un certain intervalle $x_a \leq x_i \leq x_b$. Dans l'échantillon de taille n on connaît le nombre n_a d'observations non faites dont les valeurs auraient été inférieures à x_a et le nombre n_b d'observations non faites, dont les valeurs auraient été supérieures à x_b , et on possède les $n_c = n - (n_a + n_b)$ valeurs $x_a \leq x_i \leq x_b$.

Dans ce type de troncature, n_a et n_b sont des variables aléatoires dans l'ensemble des échantillons possibles de taille n .

La méthode des moments n'est guère applicable pour déterminer les paramètres de la fonction de répartition $dF(x) = f(x)dx$.

La fonction de vraisemblance s'écrit :

$$\text{Log } \mathcal{L} = n_a \text{Log} \left[\int_{-\infty}^{x_a} f(x)dx \right] + n_b \text{Log} \left[\int_{x_b}^{+\infty} f(x)dx \right] + \sum_{x_i \geq x_a}^{x_i \leq x_b} \text{Log } f(x_i)$$

Si nous avons quelques observations dont on connaît les valeurs $x_i < a$ ou $x_i > b$, elles doivent être comptées dans les termes n_a et n_b mais non dans les termes sous Σ .

8.4.2. Il y a troncature de second type des observations (en nombre) lorsqu'on abandonne une proportion prédécidée n_a/n des plus petites valeurs observées et/ou une proportion prédécidée n_b/n des plus grandes : on ne conserve que les $n_c = n - (n_a + n_b)$ valeurs des observations $x_a \leq x_i \leq x_b$.

Dans ce type de troncature x_a et x_b sont des variables aléatoires dans l'ensemble des échantillons possibles de taille n .

La méthode des moments n'est guère applicable, et la fonction de vraisemblance s'écrit comme ci-dessus :

$$\text{Log } \mathcal{L} = n_a \text{Log} \left[\int_{-\infty}^{x_a} f(x)dx \right] + n_b \text{Log} \left[\int_{x_b}^{+\infty} f(x)dx \right] + \sum_{x_i \geq x_a}^{x_i \leq x_b} \text{Log } f(x_i)$$

On trouve ce type de troncature, par exemple, dans l'étude de durée de vie de matériels : on arrête l'expérimentation quand $n - n_b$ des appareils testés sont hors d'usage : dans ce cas la variable x est le temps, $n_a = 0$ et $x_a = 0$.

8.5. TRONQUAGE ET TRONCATURE

Nous ne nous sommes intéressés qu'à des cas simples et qui correspondent, entre autres, à l'étude des précipitations journalières, mais on pourrait imaginer des schémas plus complexes.

Dans une période donnée de n jours d'observations, le nombre de jours de pluie n'est jamais bien connu de par sa définition météorologique ; de plus les petites précipitations sont en général mal observées. Lorsqu'on cherche à ajuster une fonction de distribution continue à une série de précipitations journalières, on est conduit à prendre un paramètre de tronquage F_0 qui correspond au nombre de jours réellement sans précipitation, et un seuil de troncature x_h . On ne connaît pas le nombre de jours où on aurait dû observer une hauteur de précipitation non nulle mais intérieure à x_h ; on connaît le nombre n_j de jours où la précipitation mesurée a été supérieure ou égale à x_h et la hauteur x_j mesurée chacun de ces jours, sauf pour ceux où le pluviomètre a débordé, dont on connaît le nombre n_p et dont on sait que la précipitation a été supérieure à la capacité x_p du pluviomètre.

On choisit une fonction de distribution avec un paramètre de position connu à priori $x_0 = 0$ (borne inférieure de l'intervalle de définition de la variate). Elle s'écrit :

$$F(x) = F_0 + (1 - F_0) \int_0^x f(x) dx$$

et pour les jours de précipitations égales ou supérieures à x_h la fonction de densité s'écrit :

$$\frac{(1 - F_0) f(x)}{(1 - F_0) \int_{x_h}^{\infty} f(x) dx}$$

d'où la fonction de vraisemblance

$$\text{Log } \mathcal{L} = -n_j \text{Log} \left[\int_{x_h}^{\infty} f(x) dx \right] + n_p \text{Log} \left[\int_{x_p}^{\infty} f(x) dx \right] + \sum^{n_j - n_p} \text{Log} f(x_i)$$

Les paramètres à déterminer sont les paramètres d'échelle et de forme inclus dans la fonction de densité $f(x)$ et le paramètre de tronquage F_0 qui disparaît du système d'équations du maximum de vraisemblance. Ce dernier se détermine indirectement lorsqu'on a estimé les autres

$$F_0 = 1 - \frac{n_j}{n} \int_{x_h}^{\infty} f(x) dx$$

8.6. MÉLANGE DE DISTRIBUTIONS

Bornons-nous au cas de deux distributions, la généralisation est facile. Si $g_1(x)$ et $g_2(x)$ sont les fonctions de densité des deux populations mères (dont les variates admettent le même intervalle de définition) mélangées dans les proportions p et $(1 - p)$, la fonction de densité de la population résultante sera

$$f(x) = pg_1(x) + (1 - p)g_2(x)$$

8.6.1. La méthode des moments est d'application relativement simple lorsque les expressions mathématiques de g_1 et de g_2 ont été convenablement choisies. Le moment non centré d'ordre r du mélange est

$$m_r = pm_{1,r} + (1 - p)m_{2,r}$$

$m_{1,r}$ et $m_{2,r}$ étant les moments non centrés d'ordre r dans les distributions g_1 et g_2 . Si ces distributions comprennent chacune trois paramètres, il faudra calculer les 7 premiers moments, et les déterminations des paramètres par ce système d'équations seront difficiles à obtenir : on risque d'avoir plusieurs ensembles de solutions à départager soit par test d'adéquation, soit graphiquement.

De plus, il ne faut pas oublier que ces estimations de paramètres risquent, si la taille de l'échantillon n'est pas très importante, d'être très peu efficaces, car la variance d'un moment calculé à partir d'un échantillon croît rapidement avec son ordre.

8.6.2. Théoriquement, on peut utiliser aussi la méthode du maximum de vraisemblance, en faisant attention aux bornes. Les calculs seront encore plus difficiles que par la méthode des moments avec le même risque de pluralité des ensembles de solutions.

9. CONCLUSION

Malgré les restrictions que nous avons signalées, nous pensons que la méthode du maximum de vraisemblance doit toujours être choisie pour l'estimation des paramètres à cause de son efficacité supérieure.

Elle conduit à des estimateurs presque toujours biaisés (souvent même lorsqu'il n'y a qu'un seul paramètre à déterminer), la seule autre méthode utilisable en pratique en fait autant, et les estimateurs calculés par cette méthode des moments sont moins efficaces et probablement aussi biaisés (sauf dans certains cas d'un seul paramètre à déterminer) sinon plus.

Dans son application aux fonctions de densité continues, la méthode du maximum de vraisemblance se trouve en défaut, pour le calcul des paramètres de position (bornes de l'intervalle de définition de la variate), lorsque la fonction de densité est « en J » ou « en U ». Heureusement on peut, en général, admettre qu'alors ces bornes sont connues a priori.

Il serait intéressant, pour améliorer la méthode du maximum de vraisemblance dans le cas de petits échantillons, cas dans lequel nous nous trouvons presque toujours, de chercher à connaître, par des méthodes de Monte-Carlo, les biais des estimateurs afin de pouvoir les corriger.

9.1. REMARQUES

Les méthodes de détermination des paramètres, telles qu'elles ont été exposées, supposent que les valeurs observées sont exactement connues. Ce n'est pas le cas réel.

9.1.1. Examinons le cas où il n'y a pas d'erreurs systématiques. Quel que soit x_i observé, l'erreur à craindre sur x_i est de moyenne nulle.

Si les erreurs sont «homoscédastiques», c'est-à-dire de même variance quel que soit x_i , on peut écrire que la vraie valeur qu'on aurait dû noter se trouve dans un intervalle de confiance à tant pour cent, entre les valeurs $x_i - \Delta_1$, $x_i + \Delta_2$. En écrivant la vraisemblance de l'échantillon de taille n (7.1)

$$P_1 = \int_{x_i - \Delta_1}^{x_i + \Delta_2} f(x) dx$$

qui est peu différente, si Δ_1 et Δ_2 sont petits, de

$$(\Delta_1 + \Delta_2) f(x_i)$$

Dans les dérivations de $\text{Log } \mathcal{L}$ par rapport aux paramètres, le terme $n \text{Log} (\Delta_1 + \Delta_2)$ disparaît, les équations de la fin du paragraphe 7.1. ne sont pas modifiées.

De même, si les erreurs ne sont pas homoscédastiques mais proportionnelles à x_i , et si la vraie valeur qu'on aurait dû noter se trouve dans un intervalle de confiance $(x_i - \Delta_1 x_i)$, $(x_i + \Delta_2 x_i)$ à tant pour cent, si Δ_1 et Δ_2 sont petits, P_1 sera peu différent de

$$(\Delta_1 x_i + \Delta_2 x_i) f(x_i)$$

$\text{Log } \mathcal{L}$ devient :

$$\Sigma \text{Log } f(x_i) + \Sigma \text{Log } x_i + n \text{Log} (\Delta_1 + \Delta_2)$$

et, dans les dérivations de $\text{Log } \mathcal{L}$ par rapport aux paramètres, seuls sont conservés les termes provenant de $\Sigma \text{Log } f(x_i)$.

De même, mais cette condition n'est pas très réaliste, si les erreurs ne sont pas homoscédastiques mais proportionnelles à $f(x_i)$, si Δ_1 et Δ_2 sont petits, P_1 sera peu différent de :

$$(\Delta_1 f(x_i) + \Delta_2 f(x_i)) f(x_i)$$

$\text{Log } \mathcal{L}$ devient :

$$2 \Sigma \text{Log } f(x_i) + n \text{Log} (\Delta_1 + \Delta_2)$$

et, dans les dérivations de $\text{Log } \mathcal{L}$ par rapport aux paramètres, seuls sont conservés les termes provenant de $\Sigma \text{Log } f(x_i)$.

9.1.2. S'il y a erreurs systématiques et qu'elles soient les mêmes pour toutes les observations de l'échantillon, ces erreurs se transmettent intégralement aux paramètres d'échelle et/ou de position (suivant l'espèce d'erreur systématique) si les valeurs de ces paramètres sont inconnues a priori. Si la valeur d'un de ces deux paramètres est connue ou choisie a priori, il y a un risque que tous les paramètres déterminés par l'échantillon ne correspondent en rien au problème à résoudre.

L'effet des erreurs peut être catastrophique lorsque une partie seulement des observations est affectée d'erreur systématique de «position» et/ou d'«échelle» : par exemple étude de la répartition de hauteurs maximales de crues, les hauteurs limnimétriques ayant été lues sur des échelles différentes n'ayant pas le même calage du zéro. On peut citer comme autre exemple l'étude de la répartition des pluies journalières lorsque les précipitations ont été mesurées avec des éprouvettes tantôt convenant, tantôt ne convenant pas au pluviomètre utilisé.

BIBLIOGRAPHIE

BOREL (E.), DELTHEIL (R.), HURON (R.) - 1954 - Probabilités, erreurs. Edit. Armand Colin, Paris.

KENDALL (M.G.), STUART (A.) - 1967 - The advanced theory of statistics. Vol. 2, 2^e édit., éd. Griffin, Londres.

En particulier : chap. 4, cf. chap. 17 (Kendall), chap. 5, cf. chap. 17 (Kendall), chap. 7, cf. chap. 18 (Kendall).