Record Number:
Author, Monographic: Bobée, B.
Author Role:
Title, Monographic: Erreurs sur l'estimation des variantes hydrologiques dans le cas de la loi Pearson III
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1972
Original Publication Date:
Volume Identification:
Extent of Work: 31
Packaging Method: pages
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, Rapport de recherche
Series Volume ID: 2
Location/URL:
ISBN: 2-89146-001-4
Notes: Rapport annuel 1971-1972
Abstract: 10.00$
Call Number: R000002
Keywords: rapport/ ok/ dl

Erreurs sur l'estimation des variates hydrologiques dans le cas de la loi Pearson III

INRS-Eau
UNIVERSITE DU QUEBEC
C.P. 7500, Sainte-Foy
Québec G1V 4C7

RAPPORT SCIENTIFIQUE No 2
1972

Rapport rédigé pour
INRS-Eau

par
B. Bobée

SOMMAIRE

Erreurs sur l'estimation des variates hydrologiques
dans le cas de la loi Pearson III.

En appliquant certains résultats de la théorie statistique, on détermine l'erreur commise dans l'estimation de certaines variates (moments, coefficients de variation et d'assymétrie, événement de période de retour donnée) dans le cas de la loi Pearson III et Log Pearson III. On démontre également que la moyenne d'un échantillon d'une population qui suit une loi gamma est distribuée selon cette même loi.

Mots-clés: hydrologie, statistique, Pearson III, estimation, variates.

Bobée, B. Erreurs sur l'estimation des variates hydrologiques dans le cas de la loi Pearson III. Québec, INRS-Eau, 1972. Rapport technique no 2. 31 p.

ABSTRACT

Errors in the estimation of hydrologic variates
in the case of Pearson III law.

By the application of some results of statistical theory, we can determine the errors in the estimation of certains variates (moments, coefficient of variation, coefficient of skewness, event of given return period) in the case of Pearson III and Log Pearson III laws. In addition, we show that the mean of a sample from a gamma population is distributed following this law.

Key-words: hydrologic data, statistical models, data collection, Pearson III distribution.

Bobée, B. Erreurs sur l'estimation des variates hydrologiques dans le cas de la loi Pearson III. Québec, INRS-Eau, 1972. Rapport technique no 2. 31 p.

TABLE DES MATIERES

A1-1	Introduction
A1-2	Rappel de résultats statistiques
A1-2.1	Distribution des moments non centrés
A1-2.2	Distribution des moments centrés
A1-2.3	Distribution d'une fonction de variables aléatoires.
A1-2.4	Détermination des moments à partir de la fonction caractéristique
A1-3	Application à la loi gamma.
A1-3.1	Détermination des moments de la loi gamma.
A1-3.2	Distribution de la moyenne d'un échantillon lorsque la population mère suit une loi gamma.
A1-3.3	Erreur-type sur la variance de l'échantillon.
A1-3.4	Erreur-type sur l'écart type.
A1-3.5	Erreur-type sur le coefficient de variation de l'échantillon.
A1-3.6	Comparaison entre les erreurs-types relatives sur Cv et σ.
A1-3.7	Erreur-type sur le coefficient d'asymetrie
A1-3.8	Erreur-type sur l'estimation d'un événement ayant une période de retour donnée
A1-4	Application à la loi Log-Pearson III
A1-5	Estimateurs biaisés et non biaisés.

A1-1 INTRODUCTION

La loi de Pearson III est utilisée de plus en plus couramment pour rendre compte de phénomènes hydrologiques. Son utilisation peut se faire sous différentes formes:

- loi avec ou sans paramètre d'origine (loi gamma)
- loi log Pearson

De manière pratique, lorsque l'on considère un échantillon de taille donnée tiré d'une population qui suit une loi Pearson III, il est intéressant de connaître les erreurs-types commises lors de l'estimation de différents moments.

C'est pourquoi, après avoir rappelé les principaux résultats de statistique que nous utiliserons, nous les appliquerons à la loi Pearson III.

A1-2 RAPPELS DE RESULTATS STATISTIQUES

Considérons un échantillon comprenant N éléments x_j ($j= 1, .., N$) supposés indépendants, tirés d'une population dont:

- le moment non centré d'ordre r est μ'_r
- le moment centré par rapport à la moyenne d'ordre r est μ_r

soient:

m'_r le moment non centré d'ordre r de l'échantillon

m_r le moment centré d'ordre r par rapport à la moyenne.

on a:

$$m'_r = \frac{1}{N} \sum_{j=1}^{j=N} x_j^r \tag{1}$$

$$m_r = \frac{1}{N} \sum_{1}^{N} (x_j - m'_1)^r \tag{2}$$

Ces moments sont des estimations (déterminées à partir des valeurs de l'échantillon) des moments de la population; leur valeur est fonction de l'échantillon tiré, ces moments peuvent donc être considérés comme des variables aléatoires.

Donnons les résultats concernant les distributions de m_r et m'_r établis par Kendall (1969).

A1-2.1 DISTRIBUTION DES MOMENTS NON CENTRES (m'_r)

m'_r est distribué avec:

une moyenne $E(m'_r) = \mu'_r$ (3)

une variance $\mathrm{Var}(m'_r) = \frac{1}{N} \left[\mu'_{2r} - \mu'^2_r\right]$ (4)

une covariance $\quad Cov(m'_r, m'_q) = \frac{1}{N} \left[\mu'_{q+r} - \mu'_q \, \mu'_r \right]$ (5)

A1-2.2 DISTRIBUTION DES MOMENTS CENTRES (m_r)

Le moment centré par rapport à la moyenne, m_r , est distribué avec:

une moyenne $\quad E(m_r) \simeq \mu_r$ (6)

une variance $\quad Var(m_r) \simeq \frac{1}{N} \left[\mu_{2r} - \mu_r^2 + r^2 \mu_2 \mu_{r-1}^2 - 2r \mu_{r-1} \mu_{r+1} \right]$ (7)

on peut également montrer que la covariance de 2 moments centrés est:

$$Cov\ (m_r, m_q) \simeq \frac{1}{N} \left[\mu_{r+q} - \mu_r \mu_q + rq\ \mu_2 \mu_{r-1} \mu_{q-1} - r \mu_{r-1} \mu_{q+1} - q \mu_{r+1} \mu_{q-1} \right] \quad (8)$$

Dans le cas des moments centrés, ces formules sont des approximations à l'ordre $1/\sqrt{N}$, c'est à dire que les termes en $\frac{1}{N}, \frac{1}{N^{3/2}}$, ... etc... du développement de m_r suivant les puissances de r sont négligés.

Il est également possible d'établir au même ordre d'approximation des relations donnant la covariance entre un moment centré et un moment non centré, la plus utilisée est:

$$\mathrm{Cov}(m_1', m_r) \simeq \frac{1}{N} (\mu_{r+1} - r\mu_2 \quad \mu_{r-1}) \tag{9}$$

A1-2.3 DISTRIBUTION D'UNE FONCTION DE VARIABLES ALEATOIRES

Soit: $g(y_1, --y_i, --y_k)$ une fonction des variables aléatoires y_i $(i=1, \ldots, k)$ telles que:

$$E(y_i) = \theta_i$$

Si l'on désigne par:

θ le point de coordonnées $(\theta_1, \ldots, \theta_i, \ldots, \theta_k)$

y le point de coordonnées $(y_1, \ldots, y_i, \ldots, y_k)$

$g(y)$ est distribué avec:

une moyenne $$E\left[g(y)\right] \simeq g(\theta) \tag{10}$$

une variance

$$\mathrm{var}\left[g(y)\right] \simeq \sum_{i=1}^{k} \left[\left[\frac{\partial g(y)}{\partial y_i}\right]_{y=\theta}\right]^2 * \mathrm{var}\ yi \tag{11}$$

$$+ \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{j=1}^{k} \left[\frac{\partial g(y)}{\partial y_i}\right]_{y=\theta} \left[\frac{\partial g(y)}{\partial y_j}\right]_{y=\theta}^{*} \operatorname{cov}(y_i, y_j)$$

De même la covariance de 2 fonctions de variables aléatoires $g(y)$ et $h(y)$ est:

$$\operatorname{Cov}\left[g(y),\, h(y)\right] \simeq \sum_{i=1}^{k} \left[\frac{\partial g(y)}{\partial y_i}\right]_{y=\theta} \cdot \left[\frac{\partial h(y)}{\partial y_i}\right]_{y=\theta} \cdot \operatorname{var} x_i$$

$$+ \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} \left[\frac{\partial g(y)}{\partial y_i}\right]_{y=\theta} \left[\frac{\partial h(y)}{\partial y_j}\right]_{y=\theta} \cdot \operatorname{cov}(y_i, y_j) \tag{12}$$

A1-2.4 DETERMINATION DES MOMENTS D'UNE LOI A PARTIR DE LA FONCTION CARACTERISTIQUE DE LA DISTRIBUTION

Soit une distribution définie pour un domaine donné de x par la densité de probabilité $f(x)$

La fonction caractéristique de cette distribution est:

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx \tag{13}$$

Les moments non centrés de la distribution peuvent se déduire du développement limité de $\phi(t)$ suivant les puissances de t :

$$\phi(t) = \sum_r \frac{(it)^r}{r!} \mu'_r \qquad (14)$$

μ'_r apparaît donc comme le coefficient de $\frac{(it)^r}{r!}$ du développement

Les moments centrés par rapport à la moyenne (μ_r) sont d'autre part liés aux moments non centrés (μ'_r) par la relation:

$$\mu_r = \sum_{j=0}^{j=r} C_r^j \, (\mu'_{r-j}) \, (-\mu'_1)^j \qquad (15)$$

Il est donc possible, connaissant la fonction caractéristique, d'en déduire rapidement l'ensemble des moments μ_r et μ'_r .

Deux autres résultats importants de la fonction caractéristique peuvent être signalés.

- Si $z = \sum_{j=1}^{N} y_j$ est la somme de N variables indépendantes y_j de fonctions caractéristiques respectives $\phi_j(t)$, la fonction caractéristique de z est le produit des $\phi_j(t)$. Cette propriété résulte directement de la structure de la fonction caractéristique.

$$\phi_z(t) = \phi_1(t) * \ldots * \phi_j(t) * \ldots * \phi_N(t) \qquad (16)$$

- Si on considère la variable $u=z/p$, la fonction caractéristique de u est :

$$\phi_u(t) = \phi_z(t/p) \qquad (17)$$

A1-3 APPLICATIONS A LA LOI GAMMA

La loi gamma est la loi Pearson III sans paramètre d'origine (ou avec paramètre d'origine nul).

A1-3.1 DETERMINATION DES MOMENTS DE LA LOI GAMMA

La fonction densité de la loi gamma est:

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1} \qquad \text{avec} \quad \begin{cases} \alpha>o \\ \lambda>o \\ X\geqslant o \end{cases} \qquad (18)$$

D'après (13) la fonction caractéristique est:

$$\phi(t) = \int_0^{+\infty} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{x(it-\alpha)} dx$$

$$\phi(t) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \frac{1}{(\alpha-it)^\lambda} \Gamma(\lambda) = \frac{1}{\left(1 - \frac{it}{\alpha}\right)^\lambda} \tag{19}$$

Le développement de $\phi(t)$ suivant les puissances de t donne:

$$\phi(t) = 1 + \lambda\left[\frac{it}{\alpha}\right] + \ldots + \frac{\lambda(\lambda+1)\ldots(\lambda+r-1)}{r!}\left[\frac{it}{\alpha}\right]^r + \ldots$$

en identifiant avec la relation (14) on en déduit les moments non centrés

$$\mu'_r = \frac{\lambda(\lambda+1)\ldots(\lambda+r-1)}{\alpha^r} \tag{20}$$

Les moments centrés par rapport à la moyenne sont alors déterminés par la relation (15).

Le tableau 1 récapitule les résultats pour les moments jusqu'à l'ordre 4.

TABLEAU 1

Ordre r .	Moments non centrés (μ'_r)	Moments centrés / moyenne (μ_r)
1	λ / α	o
2	$\lambda(\lambda+1) / \alpha^2$	λ / α^2
3	$\lambda(\lambda+1)(\lambda+2) / \alpha^3$	$2\lambda / \alpha^3$
4	$\lambda(\lambda+1)(\lambda+2)(\lambda+3) / \alpha^4$	$3\lambda(\lambda+2) / \alpha^4$

On peut définir pour cette loi:

- le coefficient de variation $$C_v = \frac{\sqrt{\mu_2}}{\mu'_1} = \frac{1}{\sqrt{\lambda}} \qquad (21)$$

- le coefficient d'asymétrie $$C_s = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{2}{\sqrt{\lambda}} \qquad (22)$$

- le coefficient d'applatissement $$C_k = \frac{\mu_4}{\mu_2^2} = \frac{3(\lambda+2)}{\lambda} \qquad (23)$$

Certains auteurs (en vue d'une comparaison avec la loi normale) utilisent comme coefficient d'applatissement (C_k-3)

Ces coefficients sont liés par les relations:

$$C_s = 2\ C_v \tag{24}$$

$$C_k = 3\left[\frac{C_s^2}{2}+1\right] \tag{25}$$

$$C_k = 3\ \left[2\ C_v^2+1\right] \tag{26}$$

Ces résultats peuvent se généraliser dans le cas où le paramètre d'origine n'est pas nul (loi Pearson III).

Lorsque le paramètre d'origine n'est pas nul, la fonction de distribution est:

$$g(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)}\ e^{-\alpha(x-m)}\ (x-m)^{\lambda-1} \qquad \left\{\begin{array}{l}\alpha > o\\ \lambda > o\\ x \geqslant m\end{array}\right.$$

Pour cette loi on peut montrer que:

$$\mu'_1 = m+\frac{\lambda}{\alpha}$$

Les moments centrés par rapport à la moyenne s'expriment comme dans le tableau 1, en effet:

$$\mu_r = \int_m^{+\infty} \frac{\alpha^\lambda}{\Gamma(\lambda)}\ e^{-\alpha(x-m)}\ (x-m)^{\lambda-1}\ (x-m-\frac{\lambda}{\alpha})^r\ dx$$

Si on pose $u = x - m$

$$\mu_r = \int_0^{+\infty} \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha u} u^{\lambda-1} (u - \frac{\lambda}{\alpha})^r du$$

On retrouve donc la définition du moment centre d'ordre r de la loi gamma. Donc les coefficients C_s, C_k qui font intervenir des moments centrés, s'exprimeront de la même manière, tandis que C_v qui est lié à μ_1' s'exprimera différemment et le paramètre d'origine m interviendra.

En particulier la relation (25) reste vraie.

A1-3.2 DISTRIBUTION DE LA MOYENNE D'UN ECHANTILLON LORSQUE LA POPULATION MERE SUIT UNE LOI GAMMA

Soient x_1, .. x_j, ..x_N les N éléments indépendants de l'échantillon tiré d'une population qui suit une loi gamma.

La fonction caractéristique de x_j est d'après la relation (19)

$$\phi_j(t) = \frac{1}{\left[1 - i \frac{t}{\alpha}\right]^\lambda}$$

En utilisant la relation (16), la fonction caractéristique de $z = \sum_{1}^{N} x_j$ sera:

$$\phi_z(t) = \phi_1(t) * \phi_2(t) \ldots * \phi_N(t)$$

donc:

$$\phi_z(t) = \frac{1}{\left[1 - i\,\frac{t}{\alpha}\right]^{N\lambda}}$$

La fonction caractéristique de $m_1' = z/N$ moyenne de N valeurs de l'échantillon est d'après (17)

$$\phi_{m_1'}(t) = \phi_2\,(t/N) = \frac{1}{\left[1 - i\,\frac{t}{N\alpha}\right]^{N\lambda}}$$

Ce qui correspond à la fonction de distribution de m_1'

$$f\,(m_1') = \frac{(N\alpha)^{N\lambda}}{\Gamma\,(N\lambda)}\; e^{-N\alpha x}\; x^{N\lambda-1}$$

ou encore en posant $\alpha' = N\alpha$

$\lambda' = N\lambda$

$$f\,(m_1') = \frac{\alpha'^{\lambda'}}{\Gamma(\lambda')}\; e^{-\alpha' x}\; x^{\lambda'-1}$$

On en déduit donc que la moyenne m_1' des éléments d'un échantillon, tirés d'une population qui suit une loi gamma, suit aussi une loi gamma dont:

- la moyenne est: $\frac{\lambda'}{\alpha'} = \frac{\lambda}{\alpha}$ (= moyenne de la population mère)

- la variance est: $\frac{\lambda}{\alpha'^2} = \frac{1}{N} \frac{\lambda}{\alpha^2}$ (= $\frac{1}{N}$ variance de la population mère)

Ces résultats correspondent aux relations (3) et (4), avec la propriété supplémentaire que la moyenne est distribuée suivant une loi gamma.

Si σ est l'écart-type de la population mère (estimé par $\sqrt{m_2}$) l'erreur-type sur la moyenne de l'échantillon est:

$$E.T_{m_1'} = \frac{\sigma}{\sqrt{N}}$$

A1-3.3 ERREUR-TYPE SUR LA VARIANCE DE L'ECHANTILLON

D'après (6) et (7), la variance de l'échantillon (m_2) est distribuée avec:

- une moyenne $E(m_2) = \mu_2$

- une variance $\text{var}\ (m_2) \simeq \frac{1}{N}\left[\mu_4 - \mu_2^2\right]$

En appliquant ce résultat à la loi gamma on a

$$\text{var}\ m_2 \simeq \frac{\mu_2^2}{N}\ (C_k - 1) \qquad \text{(d'après 23)}$$

$$\text{var}\ m_2 \simeq \frac{2\mu_2^2}{N}\ (1 + 3C_v^2) \qquad \text{(d'après 26)}$$

L'erreur-type sur m_2 peut donc être estimée par

$$(E.T)_{m_2} = m_2 \sqrt{\frac{2}{N}(1 + 3\ C_v^2)}$$

A1-3.4 ERREUR-TYPE SUR L'ECART-TYPE DE L'ECHANTILLON

L'écart-type de l'échantillon est $\sqrt{m_2}$ en appliquant (11) on a:

$$\text{var}\ \sigma = \text{var}\ (m_2^{1/2}) \simeq \left[\left[\frac{\partial\ (m_2^{1/2})}{\partial\ m_2}\right]_{m_2 = \mu_2}\right]^2 * \text{var}\ m_2$$

$$\text{var}\ (m_2^{1/2}) \simeq \frac{1}{4N}\left[\frac{\mu_4 - \mu_2^2}{\mu_2}\right]$$

$$\text{var}\ \sigma \simeq \frac{\mu_2}{4N}\ (C_k - 1) \qquad \text{(d'après 23)}$$

$$\text{var}\ \sigma \simeq \frac{\mu_2}{2N}\ (1 + 3C_v^2) \qquad \text{(d'après 26)}$$

L'erreur-type sur l'écart-type de l'échantillon σ peut donc être estimée par:

$$(E.T)_\sigma = \frac{\sigma}{\sqrt{2N}} \sqrt{1 + 3\, C_v^2}$$

A1-3.5 ERREUR-TYPE SUR LE COEFFICIENT DE VARIATION DE L'ECHANTILLON

Soit C_v le coefficient de variation de l'échantillon

$$C_v = \frac{\sqrt{m_2}}{m_1'}$$

En utilisant la relation (11):

$$\mathrm{var}\,(C_v) = \left[\left[\frac{1}{2\sqrt{m_2}\; m_1'}\right]_{\substack{m_2 = \mu_2 \\ m_1' = \mu_1'}} \right]^2 * \mathrm{var}\, m_2 + \left[\left[-\frac{\sqrt{m_2}}{m_1'^2}\right]_{\substack{m_2 = \mu_2 \\ m_1' = \mu_1'}} \right]^2 * \mathrm{var}\, m_1'$$

$$-2 \left[\frac{1}{2\sqrt{m_2}\; m_1'}\right]_{\substack{m_2 = \mu_2 \\ m_1' = \mu_1'}} * \left[\frac{\sqrt{m_2}}{m_1'^2}\right]_{\substack{m_2 = \mu_2 \\ m_1' = \mu_1'}} \mathrm{Cov}\,(m_1', m_2)$$

Ce qui s'écrit

$$\text{var}\ (C_v) \simeq \frac{1}{(2\mu_1' \sqrt{\mu_2})^2} \ \text{var}\ m_2 + \frac{\mu_2}{\mu_1'^4}\ \text{var}\ m_1' - \frac{1}{\mu_1'^3}\ \text{cov}\ (m_1', m_2)$$

avec

$$\text{var}\ m_2 \simeq \frac{1}{N}(\mu_4 - \mu_2^2)$$

$$\text{var}\ m_1' = \frac{\mu_2}{N}$$

$$\text{cov}\ (m_2, m_1') \simeq \frac{\mu_3}{N} \qquad \text{(d'après 9)}$$

on obtient:

$$\text{var}\ (C_v) = \frac{C_v^2}{N}\left[\frac{\mu_4 - \mu_2^2}{4\mu_2^2} + \frac{\mu_2}{\mu_1'^2} - \frac{\mu_3}{\mu_1'\mu_2}\right]$$

en appliquant à la loi gamma, d'après (21), (22), (23):

$$\text{var}\ (C_v) = \frac{C_v^2}{N}\left[\frac{C_k - 1}{4} + C_v^2 - C_s\ C_v\right]$$

en utilisant (24) et (26) on peut exprimer cette relation

en fonction de Cv:

$$\text{var}\ (C_v) = \frac{C_v^2}{2N}\ (1 + C_v^2)$$

L'erreur-type sur le coefficient de variation peut donc être estimée par:

$$(E.T)_{C_v} = \frac{C_v}{\sqrt{2N}} \sqrt{1 + C_v^2}$$

A1-3.6 COMPARAISON ENTRE LES ERREURS-TYPES RELATIVES SUR C_v ET σ

Le rapport des erreurs-types relatives de C_v et σ est

$$\frac{(E.T)_{C_v} / C_v}{(E.T)_\sigma / \sigma} = \sqrt{\frac{1 + C_v^2}{1 + 3\, C_v^2}}$$

Dans le cas de la loi gamma ce rapport est inférieur à l'unité, cela s'explique par l'existence d'une corrélation entre σ et m_1' , en effet

$$\rho_{\sigma,\, m_1'} = \frac{\text{cov}\left[\sqrt{m_2},\, m_1'\right]}{\sqrt{\text{var}\left[\sqrt{m_2}\right] * \text{var}\,(m_1')}}$$

or:

$$\mathrm{cov}\,(\sqrt{m_2},m_1') = \left[\frac{1}{2\sqrt{m_2}}\right]_{m_2=\mu_2} \mathrm{cov}\,(m_2,m_1') \qquad \text{(d'après 12)}$$

en tenant compte de (9): $\mathrm{cov}\,(m_2,m_1') = \dfrac{\mu_3}{N}$

donc:

$$\mathrm{cov}\,(\sqrt{m_2},m_1') = \frac{\mu_3}{2N\sqrt{\mu_2}}$$

De plus:

$$\mathrm{var}\sqrt{m_2} = \frac{1}{4N}\;\frac{\mu_4 - \mu_2^2}{\mu_2}$$

$$\mathrm{var}\; m_1' = \frac{\mu_2}{N}$$

Donc en remplaçant:

$$\rho_{\sigma,m_1'} = \frac{\mu_3}{\sqrt{\mu_2}\;\sqrt{[\mu_4 - \mu_2^2]}}$$

En appliquant ce résultat à la loi gamma et en remplaçant en fonction de C_v on a:

$$\rho_{\sigma,m_1'} = C_v\sqrt{\frac{2}{1 + 3\,C_v^2}}$$

En raison de cette corrélation entre σ et m_1' , les erreurs d'échantillonnage se compenseront dans le rapport $C_v = \frac{\sigma}{m_1'}$, dont l'erreur-type relative sur Cv sera inférieure à celle commise sur σ .

A1-3.7 ERREUR-TYPE SUR LE COEFFICIENT D'ASYMETRIE

La valeur estimée du coefficient d'asymétrie est:

$$C_s = \frac{m_3}{m_2^{2/3}}$$

On peut calculer la variance de C_s en utilisant la relation (11)

$$\text{var}\ (C_s) \simeq \frac{1}{\mu_2^3}\ \text{var}\ m_3 + \frac{9}{4}\ \frac{\mu_3^2}{\mu_2^5}\ \text{var}\ m_2 - \frac{3\mu_3}{\mu_2^4}\ \text{cov}\ (m_2,m_3)$$

or:

$$\text{var}\ (m_3) \simeq \frac{1}{N}\left[\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4\right] \qquad \text{(d'après 7)}$$

$$\text{var}\ (m_2) \simeq \frac{1}{N}\left[\mu_4 - \mu_2^2\right]$$

$$\text{cov}\ (m_2,m_3) \simeq \frac{1}{N}\left[\mu_5 - 4\mu_2\mu_3\right]$$

en remplaçant il vient:

$$\text{var}\ (C_s) \simeq \frac{1}{N\mu_2^3}\left[(\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4) + \frac{9}{4}\ \frac{\mu_3^2}{\mu_2^3}\ (\mu_4 - \mu_2) - \frac{3\mu_3}{\mu_2}\ (\mu_5 - 4\mu_3\mu_2)\right]$$

Les moments μ_r de la population sont inconnus et sont estimés par m_r . Cette formule doit être utilisée prudemment en raison des différentes approximations qui ont été faites pour l'établir et de l'ordre élevé des moments qui interviennent.

Matalas et Benson (1968) conscellent d'utiliser ce résultat pour $N > 100$ et lorsque C_s n'est pas trop élevé.

Dans le cas contraire il vaut mieux utiliser le résultat établi par Fisher (1931) pour la loi normale:

$$\text{var } C_s = \frac{6N\ (N-1)}{(N-2)\ (N+1)\ (N+3)} = (E.T.)^2_{C_s}$$

Cette formule permet de déterminer si C_s est significativement différent de zéro avec un niveau de confiance donné (puisque pour la loi normale $C_s = o$).

Lorsque C_s n'est pas significativement différent de zéro, la formule précédente donne un ordre de grandeur de l'erreur-type sur C_s.

A1-3.8 ERREUR-TYPE SUR L'ESTIMATION D'UN EVENEMENT AYANT UNE PERIODE DE RETOUR DONNEE

Nous voulons déterminer l'erreur-type d'un événement X_T ayant une période de retour de T années, à partir d'un échantillon de taille N, tiré d'une population qui suit la distribution Pearson III.

X_T peut être décomposé suivant l'équation de VEN TE CHOW

$$X_T = \mu_1' + K\sqrt{\mu_2}$$

μ_1' et μ_2' étant les moments de la population sont éstimés par m_1' et m_2' , donc:

$$X_T = m_1' + K\sqrt{m_2}$$

Le facteur de fréquence K est fonction du type de loi utilisé et de la période de retour T de l'événement. Pour la loi Pearson III, Harter (1969) montre que K est parfaitement défini par la connaissance de T et du coefficient d'asymétrie (C_s)

Pour la loi gamma (Pearson III sans paramètre d'origine) T et C_v sont suffisant pour déterminer K puisque $C_s = 2\ C_v$. m_1' et m_2' sont des variables aléatoires, pour déterminer la variance de X_T on peut donc appliquer la relation (11) à l'équation de CHOW d'où:

$$\text{var } X_T = \text{var } m_1' + K^2 \text{ var}\left[\sqrt{m_2}\right] + 2K \text{ cov } \left[m_1', \sqrt{m_2}\right]$$

or:

$$\text{var } m_1' = \frac{\mu_2}{N}$$

$$\text{var}\sqrt{m_2} = \frac{1}{4N}\left[\frac{\mu_4 - \mu_2^2}{\mu_2}\right] = \frac{\mu_2}{4N}\left[C_K - 1\right]$$

$$\text{cov } \left[m_1', \sqrt{m_2}\right] = \frac{\mu_3}{2N\sqrt{\mu_2}} = \frac{\mu_2}{2N} C_s$$

donc:

$$\text{var } X_T = \frac{\mu_2}{2N} \left[2 + \frac{K^2}{2}(C_K - 1) + 2K\, C_s\right]$$

Cette relation est vraie pour toute loi. Dans le cas de la loi Pearson III et de la loi gamma, C_s et C_K ne sont pas fonction du paramètre d'origine, puisque les moments centrés pas rapport à la moyenne n'en dépendent pas, donc la relation (25) est vraie pour ces 2 lois et l'on a:

$$\text{var } X_T = \frac{\mu_2}{2N} \left[2 + K^2\left[\frac{3}{4} C_s^2 + 1\right] + 2K\, C_s\right]$$

donc:

$$(E.T.)_{X_T} \simeq \frac{\sigma}{\sqrt{2N}} \sqrt{2 + K^2 \left[3/4\, C_s^2 + 1\right] + 2K\, C_s} \qquad \text{(Loi gamma et Pearson III)}$$

Dans le cas plus particulier de la loi gamma, il est possible d'exprimer cette relation en fonction de C_v (relation 24), qui en pratique est connu avec une meilleure précision.

$$(E.T.)_{X_T} \simeq \frac{\sigma}{\sqrt{2N}} \sqrt{2 + K^2 (1 + 3C_v^2) + 4K\, C_v} \qquad \text{(Loi gamma)}$$

L'écart-type sur une caractéristique X_T ayant une période de retour de T années dépend donc pour la loi Pearson III (avec ou sans paramètre d'origine):

- la période de retour T (par l'intermédiaire de K)
- du coefficient d'asymétrie de la distribution (Cs)
- de la taille de l'échantillon (N)
- de l'écart-type de la population mère estimé par celui de l'échantillon.

Hardison (1969), a calculé $(E.T)_{X_T}$ pour la loi Pearson III, mais sa formule fait apparaître ρ_σ, m'_1 qui est ensuite déterminé par échantillonnage, or comme nous l'avons montré (A1-3.6) ρ_σ, m'_1 peut s'exprimer en fonction des moments et on arrive alors aux formules déterminées précédemment.

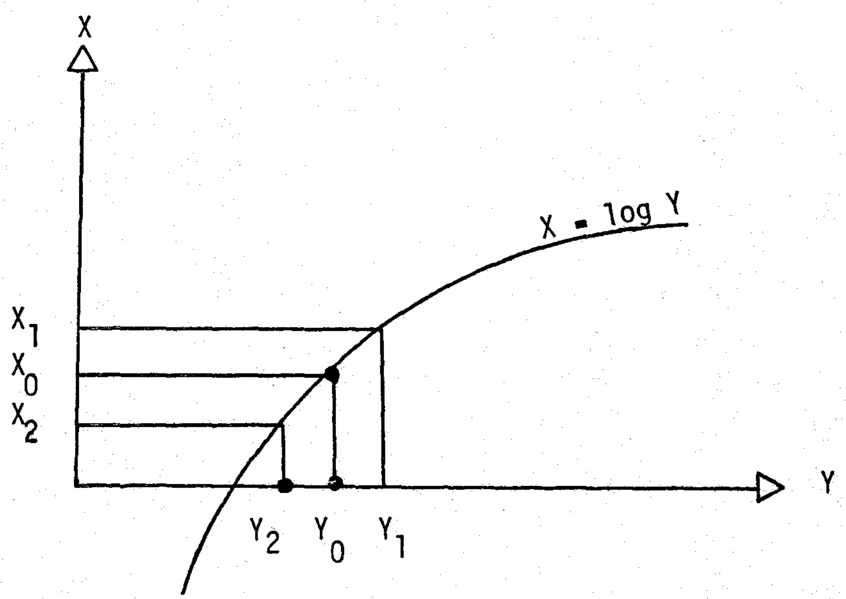
A1-4 APPLICATIONS A LA LOI LOG PEARSON III (avec ou sans paramètre d'origine)

Dans certains cas les logarithmes des événements hydrologiques sont distribués suivant une loi Pearson III (avec ou sans paramètre d'origine).

Si on désigne par $X_j = \text{Log } Y_j$, le logarithme de l'événement hydrologique Y_j , les résultats précédents peuvent donc s'appliquer, les différents moments et coefficients étant calculés sur la variable aléatoire X.

La transformation logarithmique tend en général à réduire l'asymétrie lorsque celle-ci est élevée. Hardison (1969) utilise une correspondance entre les erreurs sur X, $(E.T.)_X$, exprimées en unités logarithmiques et les erreurs relatives sur Y, $\frac{\Delta Y}{Y}$

Etablissons la relation générale donnant cette correspondance; on suppose que X=log Y suit une distribution donnée, il est alors possible de déterminer $(E.T.)_X$



on a:

$$X_1=X_0+\Delta X \qquad X_1=\log Y_1 = \log (Y_0+\Delta Y_1)$$

$$X_2=X_0-\Delta X \qquad X_2=\log Y_2 = \log (Y_0+\Delta Y_2) \qquad \Delta X,\ \Delta Y_1,\ \Delta Y_2 >0$$

$\Delta Y_1 \neq \Delta Y_2$ en raison de la non-linéarite

on peut donc écrire en identifiant les expressions de X_1 et X_2

$$X_0 + \Delta X = \log (Y_0 + \Delta Y_1) = \log Y_0 + \log (1+ \frac{\Delta Y_1}{Y_0})$$

$$X_0 - \Delta X = \log (Y_0 - \Delta Y_2) = \log Y_0 + \log (1- \frac{\Delta Y_2}{Y_0})$$

puisque $X_0 = \log Y_0$ on a:

$$\Delta X = \log (1+\frac{\Delta Y_1}{Y_0}) \qquad \text{et} \qquad -\Delta X = \log (1-\frac{\Delta Y_2}{Y_0})$$

on suppose que l'on est en logarithmes décimaux (ce qui ne nuit pas à la généralité) d'où:

$$10^{\Delta X} = 1 + \frac{\Delta Y_1}{Y_0}$$

$$10^{-\Delta X} = 1 - \frac{\Delta Y_2}{Y_0}$$

Si l'on prend comme valeur moyenne de l'erreur relative

$$\frac{\Delta Y}{Y_0} = \frac{1}{2} \left[\frac{\Delta Y_1}{Y_0} + \frac{\Delta Y_2}{Y_0}\right] \qquad \text{il vient:}$$

$$\frac{\Delta Y}{Y_0} = \frac{1}{2} \left[(10^{\Delta X} - 1) + (1 - 10^{-\Delta X})\right]$$

$$\frac{\Delta Y}{Y_0} = \frac{1}{2} \left[10^{\Delta X} - 10^{-\Delta X}\right]$$

Cette relation permet d'associer à un écart ΔX exprimé en unités logarithmiques, une erreur relative $\frac{\Delta Y}{Y_0}$ (La formule précédente se généralise aisément aux logarithmes népériens).

A1-5 ESTIMATEURS BIAISES ET NON BIAISES

m_r calculé par la formule (2) est distribué avec une moyenne $E(m_r) \simeq \mu_r$, cette relation est une approximation à l'ordre $\frac{1}{\sqrt{N}}$, elle sera d'autant plus proche de la réalité que la taille de l'échantillon est élevée.

Les estimateurs m_r des moments μ_r sont donc biaisés. Pour des événements hydrologiques, la taille de l'échantillon considéré est assez faible en général, et il vaut mieux alors utiliser des estimateurs non biaisés.

- un estimateur non biaisé de μ_2 est:

$$\hat{m}_2 = \frac{N}{(N-1)} m_2 \qquad \text{car} \qquad E(\hat{m}_2) = \mu_2$$

De même $\hat{\sigma} = \sqrt{\hat{m}_2}$

- un estimateur non biaisé de μ_3 est:

$$\hat{m}_3 = \frac{N^2}{(N-1)\ (N-2)} m_3 \qquad \text{car} \qquad E(\hat{m}_3) = \mu_3$$

- on prend également comme valeur non biaisée du coefficient d'asymétrie C_s

$$\hat{C}_s = \frac{\hat{m}_3}{\hat{m}_2^{\ 3/2}} = \frac{N^{1/2}\ (N-1)^{1/2}}{(N-2)} \frac{m_3}{m_2^{\ 3/2}}$$

Dans les différentes formules d'erreur-type, lorsque les échantillons considérés ont une taille réduite $(N < 30)$, on remplacera donc m_2, m_3, σ par $\hat{m}_2$, $\hat{m}_3$, $\hat{\sigma}$.

BIBLIOGRAPHIE

FISHER, R.A. (1931) The moments of the distribution for normal samples of measures of departure from normality Proc. Roy. Soc. London (A) 130, 16-28.

HARDISON, C.H. (1969) Accuracy of streamflow caracteristics. Geological Survey Research, D-210 - D-214.

HARTER, H.L. (1969) A new table of percentage points of the Pearson type III distribution. Technometrics Vol. II, no.1, 177-187.

KENDALL, M.G. (1969) The advanced theory of statistics. Vol. I, 3ème edition. Hafner Publishing Company.

MATALAS, N.C. & BENSON, M.A. (1968) Note on the standard error of the coefficient of skewness. Water Resources Research, Vol. 4, no.1.

Annexe A1

LISTE DES SYMBOLES

C_k	:	Coefficient d'applatissement
C_s	:	Coefficient d'asymetrie
$\hat{C}_s$	:	Valeur non biaisée du coefficient d'asymetrie
C_v	:	Coefficient de variation
C_r^j	:	Combinaison de j éléments pris parmi r.
Cov ( )	:	Covariance
E ( )	:	Espérance mathématique
E.T.	:	Erreur-type
f (x)	:	Densité de la probabilité de la variable aléatoire
g	:	fonction
h	:	fonction
K	:	Facteur de fréquence
m	:	Paramètre d'origine de la distribution Pearson III
m_r	:	Moment centré par rapport à la moyenne d'ordre r, tiré de l'échantillon
$\hat{m}_r$	:	moment centré d'ordre r, non biaisé

m'_r	:	Moment non centré d'ordre r, tiré de l'échantillon
N	:	Taille de l'échantillon
r	:	Ordre des moments
T	:	Période de retour
t	:	Variable de la fonction caractéristique d'une distribution
var ( )	:	variance
X_T	:	Evenement de période de retour T
x_j	:	Elément j de l'échantillon
y_i	:	Variable aléatoire
z	:	Variable aléatoire
α	:	Paramètre de la distribution Pearson III
$\Delta x, \Delta y$	:	Accroissement sur les variables x et y
θ_i	:	Moyenne de la variable aléatoire y_i.
λ	:	Paramètre de la distribution Pearson III
μ_r	:	Moment d'ordre r, centré par rapport à la moyenne tirée de la population
μ'_r	:	Moment non centré d'ordre r, tiré de la population
ρ	:	Coefficient de corrélation

σ	:Ecart type
$\hat{\sigma}$	:Valeur non biaisée de l'écart type.
$\phi(t)$; $\phi_j(t)$	:Fonction caractéristique d'une distribution statistique.

SELECTED WATER RESOURCES ABSTRACTS

INPUT TRANSACTION FORM

W

Title: ERRORS IN THE ESTIMATION OF HYDROLOGIC VARIATES IN THE CASE OF PEARSON III LAW (Erreurs sur l'estimation des variates hydrologiques dans le cas de la loi Pearson III),

7. Author(s)

Bobée, B.

9. Organization

Québec Université. Institut National de la Recherche Scientifique-Eau.

10. Project No.

11. Contract/Grant No.

13. Type of Report and Period Covered

12. Sponsoring Organization

15. Supplementary Notes

INRS-Eau, Technical Report No 2, 1972. 31 p.

16. Abstract

By the application of some results of statistical theory, we can determine the errors in the estimation of certains variates (moments, coefficient of variation, coefficient of skewness, event of given return period) in the case of Pearson III and Log Pearson III laws. In addition, we show that the mean of a sample from a gamma population is distributed following this law.

17a. Descriptors

*Hydrologic data, *Statistical models, Data collection.

17b. Identifiers

Pearson III distribution.

17c. COWRR Field & Group 07A

18. Availability	*19. Security Class. (Report)*	*21. No. of Pages*	**Send To:**
	20. Security Class. (Page)	*22. Price*	WATER RESOURCES SCIENTIFIC INFORMATION CENTER U.S. DEPARTMENT OF THE INTERIOR WASHINGTON, D. C. 20240

Abstractor M. Cantin | *Institution* INRS-Eau