

Fonctions de distribution en développement limité

YVES BRUNET-MORET
Ingénieur hydrologue à l'O.R.S.T.O.M.

Dans certains cas on peut connaître toutes les valeurs théoriques des moments ou des cumulants d'une variable aléatoire continue sans être capable d'en écrire la fonction de distribution.

Nous avons cherché quelles étaient les possibilités d'utiliser des développements limités suivant des polynômes orthogonaux d'Hermite, de Laguerre généralisés et de Jacobi.

Il semble que seuls les polynômes d'Hermite soient acceptables, à condition de les utiliser dans l'optique de Fisher, et il semble aussi qu'on puisse les utiliser avec les cumulants calculés d'après un échantillon.

- 1) Utilité des développements limités.
- 2) Généralités sur l'utilisation des polynômes orthogonaux.
- 3) Utilisation des polynômes d'Hermite.
- 4) Utilisation des polynômes de Laguerre généralisés.
- 5) Utilisation des polynômes de Jacobi.
- 6) Utilisation des séries de Fourier.
- 7) Utilisation sur un échantillon.
- 8) Annexe 1, polynômes orthogonaux de variables continues.
- 9) Annexe 2, moments et cumulants,
Annexe 3, listes FORTRAN.
- 10) Emploi des développements limités.

1. Utilité d'un développement limité de fonction de distribution

La loi de répartition d'une variable aléatoire continue peut être connue par la formulation mathématique de sa fonction de distribution (ou de sa fonction de densité) ou bien, entre autres, par l'ensemble des moments (ou première fonction caractéristique), par l'ensemble des cumulants (ou deuxième fonction caractéristique), ces ensembles étant spécifiques de la loi de répartition.

1.1. Il n'est pas toujours facile ni possible de remonter des moments ou des cumulants à la fonction de distribution rigoureusement exacte, mais on peut en trouver une représentation par développement suivant les moments ou les cumulants, développement que l'on peut limiter au nombre de termes nécessaires pour obtenir la précision que l'on désire dans un intervalle choisi de probabilités. Notons tout de suite que le nombre de termes devient très grand lorsque l'on veut obtenir avec une bonne précision des probabilités voisines de 0 ou de 1.

1.2. Nous donnons ci-dessous un exemple simple de cas où un développement limité suivant les cumulants pourrait être utilisé: répartition de la valeur d'un percentile, calculé en loi normale d'après un échantillon de taille n (ce n'est pas le problème de l'intervalle de confiance de la valeur du percentile).

La loi de répartition choisie à priori pour représenter la population mère est une loi normale dont les valeurs numériques des paramètres x_0 et s sont calculés d'après l'échantillon. La valeur numérique du percentile est $x = x_0 + st$ (t variable réduite normale correspondant à la probabilité du percentile).

Les répartitions de x_0 (Loi de Student) et de s (Loi gamma généralisée) sont indépendantes puisque la population mère est répartie suivant une loi normale. Donc le cumulatif d'ordre l de la répartition de x est égal à la somme du cumulatif d'ordre l de la répartition de x_0 plus t^l fois le cumulatif d'ordre l de la répartition de s .

1.3. Nous allons chercher des développements limités de fonctions de répartition suivant des polynômes orthogonaux de la variable (car la valeur du coefficient multiplicateur d'un de ces polynômes est indépendante du degré auquel on arrête le développement de leur suite) et en prenant comme fonctions de base les fonctions de densité de la loi normale, de la loi gamma incomplète de la loi bêta incomplète, pour couvrir trois cas possibles: variable continue dans l'intervalle $-\infty, +\infty$, variable continue dans un intervalle borné par une valeur finie (borne inférieure ou supérieure) et variable continue dans un intervalle borné inférieurement et supérieurement.

De plus nous examinerons les possibilités de développements en série de Fourier, qui correspondent aussi au cas d'une variable continue bornée inférieurement et supérieurement.

2. Généralités sur l'utilisation de polynômes orthogonaux dans les développements limités

2.1.

Soit $F(x)$ une fonction définie dans l'intervalle a, b de variation de x et $f(x)$ sa dérivée, continue dans cet intervalle. Nous nous proposons de trouver une représentation approchée de $f(x)$ dans l'intervalle a, b en utilisant une fonction de base $\theta(x)$ continue et positive dans l'intervalle a, b et la suite $Q_i(x)$ des polynômes orthogonaux par rapport à la fonction de base $\theta(x)$ et à l'intervalle a, b (pour la définition des polynômes orthogonaux, cf. paragraphe 9).

Remarquons qu'en déterminant une constante A par la condition

$$\int_a^b A \theta(x) dx = 1$$

la fonction $A \theta(x)$ représente la fonction de densité d'une fonction de distribution d'une variable aléatoire bornée inférieurement et supérieurement par les valeurs a et b .

2.2.

$f(x)$ étant une fonction de densité, posons

$$f(x) = A \theta(x) [A_0 Q_0 + \dots + A_i Q_i + \dots + A_n Q_n]$$

nous calculons les coefficients A_i en imposant les conditions:

$$\int_a^b f(x) Q_i(x) dx = \int_a^b A \theta(x) [A_0 Q_0 + \dots + A_n Q_n] Q_i(x) dx$$

en rappelant les propriétés des polynômes orthogonaux:

$$\int_a^b \theta(x) Q_i(x) Q_j(x) dx = 0 \quad \text{si } j \neq i$$

$$\int_a^b \theta(x) [Q_i(x)]^2 dx = h_i > 0$$

il vient quel que soit i compris entre 0 et n

$$\int_a^b f(x) Q_i(x) dx = A_i A h_i$$

$Q_i(x)$ étant un polynôme de degré i en x , on voit que le coefficient A_i est une fonction linéaire des moments jusqu'à l'ordre i inclus, et que A_i est indépendant des coefficients A_j précédents, tout autant que des suivants.

2.3.

Considérons la valeur du carré de l'erreur pour une valeur de x , carré pondéré par l'inverse de la fonction de densité qui sert à établir le développement limité (fonction de base multipliée par A)

$$\frac{\varepsilon^2(x)}{A \theta(x)} = \left[\frac{f(x)}{\sqrt{A \theta(x)}} - \sqrt{A \theta(x)} (A_0 Q_0 + \dots + A_n Q_n) \right]^2$$

dont la valeur moyenne est:

$$\eta^2 = \int_a^b \frac{f^2(x)}{A\theta(x)} dx - 2 \int_a^b f(x) (A_0 Q_0 + \dots + A_n Q_n) dx + \int_a^b A\theta(x) (A_0 Q_0 + \dots + A_n Q_n)^2 dx$$

cette valeur moyenne est minimale pour les valeurs de A_i (i de 0 à n) qui vérifient

$$\frac{\delta\eta^2}{\delta A_i} = 0$$

c'est-à-dire, en tenant compte des relations d'orthogonalité

$$A_i = \frac{1}{A h_i} \int_a^b f(x) Q_i(x) dx$$

même expression que ci-dessus (2.2). Les coefficients A_i ainsi calculés, la valeur moyenne du carré pondéré de l'erreur devient:

$$\eta^2 = \int_a^b \frac{f^2(x)}{A\theta(x)} dx - \sum_{i=0}^n A A_i^2 h_i$$

η^2 est nécessairement positif

$$\int_a^b \frac{f^2(x)}{A\theta(x)} dx \text{ également}$$

chacun des termes $A A_i^2 h_i$ aussi, donc si la somme

$$\int_a^b \frac{f^2(x)}{A\theta(x)} dx$$

n'est pas infinie, η^2 diminue lorsqu'on augmente le nombre de termes du développement limité.

On voit que, $f(x)$ étant supposé connu et représenté par une courbe en cloche, si l'on peut choisir $\theta(x)$ du même ordre de grandeur que $f(x)$ pour les faibles probabilités au dépassement et au non dépassement, le développement limité représentatif de la fonction de densité $f(x)$ permet de la calculer avec une excellente précision relative même aux extrémités. Il en sera de même pour la fonction de distribution.

Malheureusement ceci est un point de vue tout théorique, car on ne cherche à utiliser ces développements limités que dans les cas où l'on ne connaît pas la fonction de densité $f(x)$.

- 2.4.** En général l'erreur $\varepsilon(x)$ s'annule en $n + 1$ points de l'intervalle a, b et la représentation graphique de la fonction $A\theta(x) (A_0 Q_0 + \dots + A_n Q_n)$ oscille autour de la représentation graphique de la fonction $f(x)$, l'amplitude des oscillations étant d'autant plus petite que $\sqrt{A\theta(x)}$ est plus petit = mais cette amplitude croît en valeur relative lorsque $\sqrt{A\theta(x)}$ diminue.

La représentation de la fonction $\vartheta(x) = \int_a^x A \theta(x) (A_0 Q_0 + \dots + A_n Q_n) dx$

oscille également autour de la représentation de la fonction de distribution

$$F(x) = \int_a^x f(x) dx. \quad \vartheta(x) \text{ est nul lorsque } F(x) \text{ est nul, mais peut}$$

avoir un (ou plusieurs) minimums négatifs pour les valeurs de $F(x)$ proches de zéro (x proche de a). De même $\vartheta(x)$ vaut un lorsque $F(x)$ est égal à un, mais peut avoir un (ou plusieurs) maximums supérieurs à un lorsque $F(x)$ est voisin de un (x proche de a).

- 2.5. Parmi les polynômes orthogonaux, ceux de Tchebyscheff de première espèce jouissent d'une faveur particulière à cause de leur propriété spécifique (cf. 8.8.5.). Ils ne semblent cependant pas très intéressants au point de vue d'une utilisation dans un développement limité de fonction de densité, même si l'intervalle de définition de la variable admet des bornes inférieures et supérieures de valeurs finies: en effet, si la variable tend vers une de ces bornes, la fonction de base $A \theta(x)$ tend vers la valeur infinie, et même si l'erreur $\varepsilon(x)$ est très grande, la quantité $\frac{\varepsilon^2(x)}{A \theta(x)}$ tend vers zéro.

3. Utilisation des polynômes d'Hermite

Ces polynômes orthogonaux (cf. 8.6.) sont associés à la fonction de densité de la loi normale:

$$\theta(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

et à l'intervalle de variation $-\infty, +\infty$. Ils peuvent être utilisés même si l'intervalle de variation de la variable x dont on cherche une représentation de la fonction de distribution est borné d'un côté, ou des deux (car $\theta(x)$ ne sera pas nul pour $f(x)$ nul).

Le coefficient A (cf. 2.1.) est égal à 1.

Le coefficient de pondération de l'erreur $\frac{1}{\sqrt{\theta(u)}}$ (cf. 2.3.) a pour valeur minimale 1,58 pour $u = 0$. Lorsque $|u|$ est assez grand, après avoir posé $\frac{1}{T} = \int_{|u|}^{\infty} \theta(u) du$ (T temps de recurrence associé à la valeur $|u|$ dans la loi normale), $\theta(u)$ tend vers $\frac{|u|}{T}$ lorsque $|u|$ augmente et le coefficient de pondération tend vers $\sqrt{\frac{T}{|u|}}$. Pour $T = 1\,000$, $|u| = 3,09$ il vaut environ 18,0.

- 3.2. Nous supposons connus les cumulants d'ordre 1 à n de la variable x

k_1 moyenne

k_2 variance

...

soit directement, soit par les moments $m_1 m_2 \dots m_n$. Après avoir posé:

$$u = \frac{x - k_1}{\sqrt{k_2}}$$

Nous écrivons le développement limité de la fonction de distribution sous la forme:

$$F(x) = \int_{-\infty}^u \theta(u) du - \theta(u) \sum_{i=1}^{n-1} A_i H_i(u)$$

$H_i(u)$ étant le polynome d'Hermite de degré i
d'où le développement limité de la fonction de densité:

$$f(x) = \theta(u) + u \theta(u) \sum_{i=1}^{n-1} A_i H_i(u) - \theta(u) \sum_{i=1}^{n-1} A_i H_i'(u)$$

qui devient en utilisant les relations de 8.6.3.

$$f(x) = \theta(u) + \theta(u) \sum_{i=1}^{n-1} A_i H_{i+1}(u)$$

Après avoir multiplié les deux membres de cette équation par $H_{i+1}(u)$
(i de 1 à $n-1$) on somme de $-\infty$ à $+\infty$

$$\int_{-\infty}^{+\infty} f(u) H_{i+1}(u) du = \int_{-\infty}^{+\infty} A_i [H_{i+1}(u)]^2 du = (i+1)! A_i$$

ce qui permet de calculer les coefficients A_i car

$$\int_{-\infty}^{+\infty} f(u) H_{i+1}(u) du = (i+1)! \sum_{J=0}^N (-1)^J \frac{\mu_{i+1-2J}}{J! (i+1-2J)! 2^J}$$

avec $N =$ partie entière de $\frac{i+1}{2}$

les μ étant les moments de la variable u , c'est-à-dire les moments centrés de la
(variable x divisée par la racine carrée de sa variance):

$$A_i = \sum_{J=0}^N (-1)^J \frac{\mu_{i+1-2J}}{J! (i+1-2J)! 2^J}$$

d'où

$$\begin{aligned} A_1 &= 0 \\ A_2 &= \mu_3/3! \\ A_3 &= (\mu_4 - 3)/4! \dots \text{etc.} \dots \end{aligned}$$

Ce développement limité est connu sous le nom de Gram-Charlier.

On peut écrire les coefficients A_i sous une forme souvent plus pratique (Edgeworth):

les K_i étant les cumulants d'ordre i de la variable x ; les cumulants de la variable u sont:

$$K'_1 = 0 \quad K'_2 = 1 \quad K'_3 = \frac{K_3}{K_2^{3/2}} \quad \dots \quad K'_n = \frac{K_n}{K_2^{n/2}}$$

et

$$A_{i-1} = \sum \frac{1}{\pi J!} \pi \left(\frac{K'_1}{1!} \right)^J$$

où

$$3 \leq i \leq n, \quad J \geq 0 \quad \text{et} \quad \sum l_J = i$$

La sous-routine SPARKE calcule ces coefficients A d'après les cumulants, en les nommant S, dont l'indice IR correspond à $i + 1$.

En fait les résultats obtenus avec ce type de développement limité sont souvent décevants: (cf. 3.5.), l'emploi de p termes n'assure pas une meilleure approximation que l'emploi de $p - 1$ termes, pour toutes les fréquences. Les résultats ne sont acceptables que si les cumulants K'_3 et K'_4 (qui sont les coefficients d'asymétrie et d'aplatissement de la distribution de la variable x) sont assez voisins de zéro (distribution proche de la normale) et si les expressions $\frac{K'_i}{i!}$ tendent rapidement vers zéro lorsque i croît.

3.3. FISCHER et MAC-CORNISH ont considéré un cas particulier très intéressant: soit une variable aléatoire dont on connaît la distribution (ou les moments, ou les cumulants, ce qui revient au même), nous connaissons donc les cumulants de la distribution de la moyenne d'un échantillon de m de ces variables aléatoires indépendantes:

L_i étant le cumulants d'ordre i de la variable aléatoire

le cumulants de même ordre de la moyenne est

$$K_i = L_i/m^{i-1} \quad \text{cad} \quad K_1 = L_1 \quad K_2 = L_2/m \quad \dots$$

Les cumulants de la variable réduite correspondant à la moyenne sont:

$$K'_i = \frac{k_i}{K_2^{i/2}} = \frac{L_i}{L_2^{i/2}} \cdot \frac{1}{m^{i/2 - 1}}$$

FISCHER ne conserve, dans le développement limité, que les termes en $\left(\frac{1}{\sqrt{m}} \right)$ de degré inférieur ou égal à r. Les coefficients A sont alors calculés par

$$A_{i-1} = \sum \frac{1}{\pi J!} \pi \left(\frac{K'_1}{1!} \right)^J$$

où:

$$3 \leq l \leq i, \quad J \geq 0, \quad \sum l_j = i \quad \text{et} \quad \sum J \left(\frac{l}{2} - 1 \right) \leq \frac{r}{2}$$

et le dernier terme à calculer est A_{n-1} où $n-1 = 3r$.

La sous-routine SPARKF calcule ces coefficients A d'après les cumulants, en les nommant S, dont l'indice IR correspond à $i+1$. Le nombre L de coefficients calculés doit être pris multiple de 3: $L = 3r$ pour négliger tous les termes contenant le facteur à une puissance supérieure à r.

3.4. Il faut être très prudent dans l'utilisation de ces développements limités = on ne peut savoir, a priori, s'ils sont convergents pour toutes les valeurs de la variable puisqu'on ne connaît pas $f(x)$.

Prenons un cas très particulier = il ne semble pas possible de trouver un développement limité du genre ci-dessus pour représenter une loi de Gumbel. Pourtant les

expressions $\frac{K'_i}{i!}$ tendent vers $\frac{1}{i} \left(\frac{\sqrt{6}}{\pi} \right)^i$ c'est-à-dire vers zéro lorsque i croît. Mais

(cf. 2.3.), le rapport $\frac{f^2(x)}{\theta(x)}$

où:

$$f^2(x) = e^{-2x} e^{-2e^{-x}} \quad \text{et} \quad \theta(x) = \frac{\sqrt{3}}{\pi \sqrt{\pi}} e^{-3 \frac{(x-c)^2}{\pi^2}}$$

tend vers $+\infty$ si x tend vers $+\infty$ et vers zéro si x tend vers $-\infty$.

Dans le cas des lois gamma incomplètes, les expressions $\frac{K'_i}{i!}$ valent $\frac{1}{i} \frac{1}{\gamma^{\frac{i}{2}-1}}$ c'est-à-dire tendent vers zéro (si $\gamma > 1$) lorsque i croît. Le rapport $\frac{f^2(x)}{\theta(x)}$

où

$$f^2(x) = \left[\frac{1}{\Gamma(\gamma)} \right]^2 x^{2\gamma-2} e^{-2x} \quad \text{et} \quad \theta(x) = \frac{1}{\sqrt{2\pi\gamma}} e^{-\frac{(x-\gamma)^2}{2\gamma}}$$

est nul pour $x < 0$, tend vers $+\infty$ en même temps que x. Mais la somme $\int_{-\infty}^{+\infty} \frac{f^2(x)}{\theta(x)} dx$

est limitée et on peut trouver des développements limités acceptables ($\gamma > 1$).

Il faut donc essayer, dans le cas général, le développement limité avec différents nombres de termes pour vérifier s'il est vraiment convergent à l'intérieur de la gamme intéressante de probabilités.

3.5. Le calcul à l'ordinateur de la fonction de probabilité en développement limité fait l'objet du sous-programme FDEVS qui traduit la première équation du paragraphe 3.2. et s'utilise avec un des sous-programmes SPARKE ou SPARKF pour le calcul des coefficients multiplicateurs des polynômes d'Hermite.

L'inversion de cette fonction est possible par la méthode du pas (sous-programme VDEVS). Il est évident qu'il ne faut pas chercher une grosse précision sur l'inconnue, et surtout pas vers les faibles ou fortes probabilités.

3.6. A titre d'exemple nous donnons dans le tableau ci-dessous des résultats de calcul de fréquences d'après les développements limités du type Gram-Charlier-Edgeworth et du type Fisher Mac Cornish. La fonction de distribution mise en développement limité est celle d'une loi gamma incomplète de paramètres de forme 12, d'échelle 1, et de position zéro. Trois fréquences ont été choisies 0,0001, 0,5 et 0,9999. Pour chacune de ces fréquences la variable a été calculée en simple précision, par inversion de la loi gamma incomplète, et ces valeurs de la variable ont été utilisées pour calculer les fréquences en développements limités.

COMPARAISON DES FRÉQUENCES
CALCULÉES D'APRÈS LES DÉVELOPPEMENTS LIMITÉS
DE GRAM CHARLIER EDGEWORTH ET DE FISHER MAC CORNISH

Colonne 1 : Nombre de termes retenus
Colonnes 2 : Fréquences Gram Charlier Edgeworth (avec SPARKE)
Colonnes 3 : Fréquences Fischer Mac Cornish (avec SPARKF)

(1)	Fréquence choisie 0,00010		Fréquence choisie 0,50000		Fréquence choisie 0,99990	
	(2)	(3)	(2)	(3)	(2)	(3)
4	0,0001700		0,4973589		0,9999927	
5	-0,0004715		0,4917319		0,9999893	
6	-0,0021214	-0,0013030	0,4950120	0,4999830	0,9999722	0,9999791
7	0,0007177		0,5082853		0,9999469	
8	-0,0003875		0,5053096		0,9999185	
9	-0,0022449	-0,0000356	0,4900964	0,5000042	0,9998765	0,9999370
10	0,0017592		0,4959351		0,9998350	
11	0,0006075		0,5184712		0,9998184	
12	-0,0044518	0,0002813	0,5084464	0,5000033	0,9998496	0,9998879
13	0,0029662		0,4674008		0,9999437	
14	0,0052186		0,4881638		1,0000677	
15	-0,0117497	0,0002634	0,5688115	0,5000000	1,0001221	0,9998800
16	0,0004008		0,5217409		0,9999667	
17	0,0264157		0,3429673		0,9995438	
18	-0,0001923	0,0001923	0,4593194	0,4999999	0,9991053	0,9999091
19	-0,0387353		0,8913907		0,9993096	
20	0,1091344		0,5794570		1,0008087	
21	-0,0017678	0,0001411	0,5541440	0,4999999	1,0030565	0,9999093
22	-0,8033078		0,3439801		1,0030804	
23	0,2972319		3,5453329		0,9967170	
24	0,6245564	0,0001142	0,7822817	0,4999999	0,9850333	0,9998887
25	-1,5125837		-8,8882132		0,9832460	
26	-0,5622056		0,1473371		1,0168123	
27	5,4998398	0,0001027	31,2335205	0,4999999	1,0829020	0,9999019
28	-3,2425900		-0,0425253		1,0888004	
29	-16,1068573		-105,9105682		0,8719512	
30	25,5079346	0,0000987	8,2659016	0,4999999	0,4669241	9,0999070

Il s'introduit dans ces calculs une erreur qui n'est peut-être pas négligeable lorsque le nombre de termes retenus est grand (et donc lorsque les ordres des derniers polynômes d'Hermite sont grands) car la variable n'est introduite qu'en simple précision, alors que les coefficients des polynômes d'Hermite le sont en double précision.

Il est cependant manifeste d'une part, que les développements limités du type Gram-Charlier-Edgeworth divergent: les probabilités calculées sont de moins en moins exactes au fur et à mesure que le nombre de termes retenus augmente, et d'autre part que les développements limités du type Fisher-Mac Cornish convergent d'une façon à peu près régulière.

4. Utilisation des polynômes de Laguerre généralisés

4.1.

Ces polynômes orthogonaux (cf. 8.7) sont associés à la fonction de densité de la loi gamma incomplète:

$$\theta(u) = \frac{1}{\Gamma(\gamma)} u^{\gamma-1} e^{-u}$$

et à l'intervalle de variation $0, +\infty$. Il semble qu'on ait avantage à ne les utiliser que si l'intervalle de variation de la variable aléatoire x dont on cherche une représentation de la fonction de distribution est borné inférieurement (ou supérieurement) par une valeur finie.

Le coefficient A (cf. 2.) est égal à 1.

Le coefficient de pondération de l'erreur $\frac{1}{\sqrt{\theta(u)}}$ (cf. 2.3) dépend évidemment de γ .

Il tend vers zéro lorsque, γ étant inférieur à 1, u tend vers zéro, ce qui montre qu'on n'a pas intérêt à choisir une valeur de γ inférieure à 1 lorsqu'on est sûr que la variable aléatoire a une fonction de distribution unimodale. Lorsque u croît, et en posant

$$\frac{1}{T} = \int_u^{\infty} \theta(u) du \quad (T \text{ temps de recurrence associé à la valeur } u) \quad T \text{ tend vers } \frac{1}{\theta(u)}$$

le coefficient de pondération tend vers \sqrt{T} .

4.2. Nous supposons connus les cumulants (et les moments) d'ordres 1 à n de la variable x , bornée inférieurement.

4.2.1.

Nous pouvons chercher à déterminer les paramètres de forme γ d'échelle s et de position x_0 d'une distribution gamma incomplète adaptée aux trois premiers cumulants:

$$x_0 = 2 \frac{K_2^2}{K_3} - K_1$$

$$s = K_3 / 2 K_2$$

$$\gamma = 4 K_2^3 / K_3^2$$

(s est du même signe que K_3 , qui doit être positif pour que la borne soit inférieure).

4.2.2.

Si l'on connaît, ou si l'on peut choisir à priori la valeur de la borne (inférieure) x_0 de l'intervalle de variation de x , on peut calculer s et γ par les deux premiers cumulants:

$$s = K_2 / (K_1 - x_0)$$

$$\gamma = (K_1 - x_0)^2 / K_2$$

4.3. Après avoir posé $u = \frac{x - x_0}{s}$ nous écrivons le développement limité de la fonction de distribution sous la forme:

$$F(x) = \int_0^u \theta(u) du + u \theta(u) \sum_{i=1}^n B_i L_i(u)$$

$L_i(u)$ étant le polynome de Laguerre généralisé de degré i d'où le développement limité de la fonction de densité:

$$\begin{aligned} f(x) &= \theta(u) + u \theta(u) \sum_1^n B_i L'_i(u) \\ &+ \gamma \theta(u) \sum_1^n B_i L_i(u) \\ &- u \theta(u) \sum_1^n B_i L_i(u) \end{aligned}$$

qui devient en utilisant les relations de 8.7.3.

$$f(x) = \theta(u) + \theta(u) \sum_1^n (i B_i + B_{i-1}) L_i(u) \quad \text{avec} \quad B_0 = 0$$

Après avoir multiplié les deux membres de cette expression par $L_i(u)$ (i de 1 à n) on somme de 0 à $+\infty$

$$\int_0^\infty f(u) L_i(u) du = \int_0^\infty (i B_i + B_{i-1}) [L_i(u)]^2 du = (i B_i + B_{i-1}) i! \frac{\Gamma(\gamma + i)}{\Gamma(\gamma)}$$

ce qui permet de calculer les coefficients B_i car

$$\int_0^\infty f(u) L_i(u) du = i! \sum_{j=0}^i \frac{(-1)^j}{j! (i-j)!} \frac{\Gamma(\gamma + i)}{\Gamma(\gamma + i - j)} m_{i-j}$$

les m_{i-j} étant les moments d'ordre $i-j$ de la variable u que l'on peut calculer d'après ceux de la variable x ($m_0 = 1$).

4.4. Les coefficients B_i peuvent se calculer par recurrence:

$$iB_i + B_{i-1} = \sum_{j=0}^i \frac{(-1)^j}{j!(i-j)!} \frac{\Gamma(\gamma)}{\Gamma(\gamma+i-j)} m_{i-j}$$

d'où l'on peut tirer:

$$B_i = \frac{(-1)^i}{i!} \sum_{l=1}^i \frac{1}{l} + \frac{1}{i!} \sum_{l=1}^i (-1)^{i-l} \frac{\Gamma(\gamma)}{\Gamma(\gamma+l)} \frac{m_l}{l!} \sum_{j=l}^i \frac{(j-1)!}{(j-l)!}$$

ou m_1 est le moment d'ordre 1 de la variable $u = \frac{x-x_0}{s}$

Les premiers coefficients B_i sont nuls en nombre égal au nombre de paramètres que l'on a pu ajuster de la loi gamma incomplète (cf. 4.2.).

4.5. Cas particulier: polynomes de Laguerre proprement dits $\gamma = 1$. Nous supposons connue à l'avance la valeur de la borne x_0 . On peut calculer le paramètre d'échelle nécessaire:

$$s = K_1 - x_0 \quad \text{d'où } m_1 = 1$$

$$B_1 \text{ est nul} \quad B_2 = -\frac{1}{4} + \frac{1}{16} m_2 \text{ etc....}$$

4.6. On ne peut appliquer aux développements limités suivant des polynomes de Laguerre généralisés l'idée de FISHER (cf. 3.3.): car en exprimant les moments en fonction des cumulants pour ne garder que les termes de degré inférieur ou égal à r en $\frac{1}{n}$, on ne peut pas trouver de limite à la suite des coefficients B_i , alors que leur nombre est limité dans le cas des polynomes d'Hermite.

C'est pour cette raison que nous n'avons pas cherché à établir de programme pour l'utilisation des polynomes de Laguerre généralisés.

Il faut sûrement être très prudent dans leur application. Un des points dangereux semble être le choix ou le calcul de la valeur de la borne x_0 . Si par malheur $f(x)$ n'est pas nul pour $x < x_0$, $\theta(u)$ l'est et le rapport $\frac{f^2(x)}{\theta(u)}$ (cf. 2.3.) n'a plus de sens.

5. Utilisation des polynomes de Jacobi

5.1.

Ces polynomes orthogonaux (cf. 8.8.) sont associés à une fonction de base qui est la fonction de densité de la loi bêta incomplète multipliée par $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ et à l'intervalle de variation 0,1. Il semble qu'on ait avantage à ne les utiliser que si l'intervalle de variation de la variable aléatoire x dont on cherche une représentation de la fonction de distribution admet des bornes inférieures et supérieures de valeurs finies.

Le coefficient A. (cf. 2.1.) est $\frac{1}{B(p, q)}$

Le coefficient de pondération de l'erreur $\frac{1}{\sqrt{\theta(u)}}$ (cf. 2.3.) dépend évidemment de p et de q. Il tend vers zéro lorsque p étant inférieur à 1, u tend vers zéro, et lorsque, q étant inférieur à un, u tend vers 1; ce qui montre qu'on n'a pas intérêt à choisir des valeurs de p et de q inférieures à 1 lorsqu'on est sûr que la fonction de densité de la variable aléatoire est unimodale en cloche (cf. 2.5.).

5.2. Nous supposons connus les cumulants (et les moments) d'ordres 1 à n de la variable x, bornée inférieurement et supérieurement.

5.2.1.

Nous pouvons chercher à déterminer les paramètres de forme p et q et de position x_0 et x_1 d'une distribution bêta incomplète adaptée aux quatre premiers cumulants:

$$s = x_1 - x_0$$

$$K_1 = x_0 + s \frac{p}{p+q} \quad K_2 = s^2 \frac{p q}{(p+q)^2 (p+q+1)}$$

$$K_3 = -s^3 \frac{2 p q (p-q)}{(p+q)^3 (p+q+1) (p+q+2)}$$

$$K_4 = s^4 \frac{6 p q [(p+q)^2 (p+q+1) - p q [5(p+q) + 6]]}{(p+q)^4 (p+q+1)^2 (p+q+2) (p+q+3)}$$

d'où:

$$\frac{K_3^2}{K_2^3} + 4 = 4 \frac{(p+q)^2 (p+q+1 + p q)}{p q (p+q+2)^2}$$

$$\frac{K_4}{K_2^2} + 6 = 6 \frac{(p+q)^2 (p+q+1 + p q)}{p q (p+q+2) (p+q+3)}$$

$$p+q = 6 \frac{K_2 K_4 + 2 K_2^3 - K_3^2}{3 K_2^2 - 2 K_2 K_4}$$

première condition $p+q > 0$.

puis:

$$p q = 4 (p+q)^2 (p+q+1) / \left[\frac{K_3^2}{K_2^3} (p+q)^2 + 4 \left(\frac{K_3^2}{K_2^3} + 4 \right) (p+q+1) \right]$$

ou:

$$p q = 6 (p+q)^2 (p+q+1) / \left[\frac{K_4}{K_2^2} (p+q)^2 + \left(\frac{K_4}{K_2^2} + 6 \right) (5(p+q) + 6) \right]$$

p q étant positif si p + q l'est

deuxième condition $(p + q)^2 - 4 p q > 0$ pour que p et q soient les racines de $Z^2 - (p + q) Z + p q = 0$
 s, x_0 et x_1 , sont calculés par :

$$s = (p + q) \sqrt{\frac{K_2 (p + q + 1)}{p q}}$$

$$x_0 = K_1 - s \frac{p}{p + q}$$

$$x_1 = x_0 + s$$

5.2.2.

Si on ne peut calculer p et q , les conditions signalées ci-dessus n'étant pas remplies, on peut choisir p et q de façon que

$$\frac{4 (p - q)^2 (p + q + 1)}{p q (p + q + 2)^2} = \frac{K_3^2}{K_2^3}$$

et ensuite calculer s, x_0 et x_1 comme ci-dessus.

5.2.3.

Si l'on connaît, ou si l'on peut choisir les valeurs des bornes x_0 et x_1 , on peut calculer p et q .

$$p + q = \frac{(x_1 - K_1) (K_1 - x_0) K_2}{K_2}$$

avec comme condition $(x_1 - k_1) (K_1 - x_0) > K_2$
d'où :

$$p = \frac{K_1 - x_0}{s} (p + q)$$

$$q = \frac{x_1 - K_1}{s} (p + q)$$

5.2.4.

Si l'on connaît une seule des deux bornes, on peut calculer l'autre et les paramètres de forme à l'aide des trois premiers cumulants.

5.3. Nous écrivons le développement limité de la fonction de distribution sous la forme, après avoir posé $u = \frac{x - x_0}{s}$

$$F(x) = \int_0^u \frac{\theta(u)}{B(p, q)} du + \frac{\theta(u)}{B(p + q)} u (1 - u) \sum_{i=1}^n B_i G_i(u)$$

$G_i(u)$ étant le polynome de Jacobi de degré i

d'où le développement limité de la fonction de densité

$$f(x) = \frac{\theta(u)}{B(p,q)} + \frac{\theta(u)}{B(p,q)} P \sum_1^n B_i G_i(u) - \frac{\theta(u)}{B(p,q)} (p+q) \sum_1^n u B_i G_i(u) + \frac{\theta(u)}{B(p,q)} \sum_1^n u(1-u) B_i G_i'(u)$$

qui devient en utilisant les relations de 8.8.3.

$$f(x) = \frac{\theta(u)}{B(p,q)} + \frac{\theta(u)}{B(p,q)} \sum_1^n \left[-(i+p+q-1) B_{i-1} + \frac{i(p-q)(i+p+q-1)}{(2i+p+q-2)(2i+p+q)} B_i + \frac{i(i+1)(i+p)(i+q)(i+p+q-1)}{(2i+p+q-1)(2i+p+q)^2(2i+p+q+1)} B_{i+1} \right] G_i(u)$$

avec $B_0 = 0$

après avoir multiplié les deux membres de cette expression par $G_i(u)$ (i de 1 à n , avec $B_{n+1} = 0$), on somme de 0 à 1

$$\int_0^1 f(u) G_i(u) du = \frac{\Gamma(i+p) i!}{\Gamma(2i+p+q-1)} \sum_{j=0}^i \frac{(-1)^j}{j! (i-j)!} \frac{\Gamma(2i+p+q-j-1)}{\Gamma(i+p-j)} m_{i-j} = \frac{i!}{B(p,q)} \frac{\Gamma(i+p) \Gamma(i+q) \Gamma(i+p+q-1)}{\Gamma(2i+p+q) \Gamma(2i+p+q-1)} \times (\text{coefficient en } B_{i-1}, B_i, B_{i+1} \text{ ci-dessus de } G_i(u))$$

les m_{i-j} étant les moments d'ordre $(i-j)$ de la variable u que l'on peut calculer d'après ceux, connus, de la variable x ($m_0 = 1$).

On trouve ainsi les valeurs des coefficients B_i . Les premiers sont nuls en nombre égal au nombre de paramètres que l'on a pu ajuster de la loi bêta incomplète (cf. 5.2.).

Cela ne semble pas très maniable, même si l'on peut faire $B_1 = 0$. Nous allons envisager trois cas particuliers.

5.4. Polynomes de Tchebyscheff de première espèce $p = q = 1/2$. Les expressions se simplifient et en tenant compte de la relation

$$\Gamma(2i+1) = \frac{1}{\sqrt{\pi}} 2^{2i} \Gamma(i + 1/2) \Gamma(i+1)$$

il vient :

$$\frac{i!}{\sqrt{\pi} 2^{2i}} (-B_{i-1} + \frac{1}{16} B_{i+1}) = \sum_{j=0}^i \frac{(-1)^j i!}{j! (i-j)!} \frac{\Gamma(2i-j)}{\Gamma(i + \frac{1}{2} - j)} m_{i-j}$$

que l'on peut écrire:

$$\frac{i}{2^{2i}} (-B_{i-1} + \frac{1}{16} B_{i+1}) = \sum_{j=0}^i \frac{(-1)^j i!}{j! (i-j)!} \frac{\Gamma(1/2)}{\Gamma(i)} \frac{\Gamma(i+i-j)}{\Gamma(\frac{1}{2} + i - j)} m_{i-j}$$

Les bornes x_0 et x_1 , de l'intervalle de variation de la variable x étant connues à l'avance, et les m_{i-j} étant les moments d'ordre $i-j$ de la variable réduite $u = \frac{x - x_0}{x_1 - x_0}$ il vient

$$\begin{aligned} B_1 &= 0 \\ \frac{1}{64} B_2 &= 2 m_1 - 1 \\ \frac{1}{128} B_3 &= 8 m_2 - 8 m_1 + 1 \quad \text{etc.} \end{aligned}$$

5.5. Polynomes de Tchebyscheff de seconde espèce

Polynomes de TCHEBYSCHEFF de seconde espèce

$p = q = 3/2$, nous obtenons:

$$\frac{2i!}{\sqrt{\pi} 2^{2i}} \left[-(i+2) B_{i-1} + \frac{i}{16} B_{i+1} \right] = \sum_{j=0}^i \frac{(-1)^j i!}{j! (i-j)!} \frac{\Gamma(2i+2-j)}{\Gamma(i+3/2-j)} m_{i-j}$$

que l'on peut écrire:

$$\frac{1}{2^{2i}} \left[-\frac{i+2}{i+1} B_{i-1} + \frac{i}{16(i+1)} B_{i+1} \right] = \sum_{j=0}^i \frac{(-1)^j i!}{j! (i-j)!} \frac{\Gamma(3/2)}{\Gamma(i+2)} \frac{\Gamma(i+2+i-j)}{\Gamma(3/2+i-j)} m_{i-j}$$

Les bornes x_0 et x_1 de l'intervalle de variation de la variable x étant connues à l'avance et les m_{i-j} étant les moments d'ordre $i-j$ de la variable réduite $u = \frac{x - x_0}{x_1 - x_0}$ il vient:

$$\begin{aligned} B_1 &= 0 \\ \frac{1}{128} B_2 &= 2 m_1 - 1 \\ \frac{1}{128} B_3 &= 16 m_2 - 16 m_1 + 3 \quad \text{etc.} \end{aligned}$$

5.6. Polynomes de Legendre

Polynomes de Legendre

$p = q = 1$ nous obtenons:

$$\frac{i!(i+1)!}{(2i+1)!} \left[-B_{i-1} + \frac{i(i+1)}{4(2i+1)(2i+3)} B_{i+1} \right] = \sum_{j=0}^i \frac{(-1)^j}{j! (i-j)!} \frac{\Gamma(2i+1-j)}{\Gamma(i+1-j)} m_{i-j}$$

Les bornes x_0 et x_1 de l'intervalle de variation de la variable x étant connues à l'avance et les m_{i-j} étant les moments d'ordre $i - j$ de la variable réduite $u = \frac{x - x_0}{x_1 - x_0}$ il vient :

$$B_1 = 0.$$

$$\frac{1}{90} B_2 = 2 m_1 - 1$$

$$\frac{3}{100} B_3 = 6 m_2 - 6 m_1 + 1 \quad \text{etc.}$$

5.7. On ne peut appliquer aux développements limités suivant des polynomes de Jacobi, l'idée de FISHER (cf. 3.3.): car, en exprimant les moments en fonction des cumulants pour ne garder que les termes de degré inférieur ou égal à r en $\frac{1}{n}$, on ne peut trouver de borne supérieure à la suite des coefficients B_i (alors que le nombre de ces coefficients est limité dans le cas des polynomes d'Hermite).

Cette remarque explique que nous n'ayons pas rédigé de programme pour l'utilisation des polynomes de Jacobi.

Il faut sûrement être très prudent dans leur application. Il semble nécessaire de connaître à priori les valeurs des bornes de l'intervalle de variation de x . Un mauvais choix de ces bornes ou leur calcul indiqué en 5.2. conduit sûrement à de très mauvaises estimations des probabilités fortes ou faibles.

6. Utilisation de séries de Fourier

L'intervalle de variation de la variable aléatoire étant 0,1, on peut écrire la fonction de densité sous la forme:

$$f(x) = A_0 + \sum_{i=1}^n (A_i \cos 2 i \pi x + B_i \sin 2 i \pi x)$$

6.1. Développement en série de Fourier

et la fonction de distribution:

$$F(x) = A_0 x + \frac{1}{2\pi} \sum_{i=1}^n \left(\frac{A_i}{i} \sin 2 i \pi x - \frac{B_i}{i} \cos 2 i \pi x \right)$$

Il y a une certaine analogie entre le terme de rang i du développement en série de Fourier et le polynome d'ordre $2 i$ de Tchebyscheff de première espèce: ils s'annulent pour $2 i$ valeurs de la variable, et les écarts absolus de leurs maximums ou de leurs minimums à leur moyenne sont égaux entre eux.

6.2. Le calcul des coefficients A_i et B_i n'est pas simple

Nous pouvons prendre comme premières conditions:

$$F(0) = 0$$

$$F(1) = 1$$

ce qui conduit à:

$$A_0 = 1 \quad \sum \frac{B_i}{i} = 0$$

en partant de:

$$\int_0^1 A_i x^m \cos 2 i \pi x dx = A_i \frac{m}{(2 i \pi)^2} - \frac{m(m-1)}{(2 i \pi)^2} \int_0^1 A_i x^{m-2} \cos 2 i \pi x dx$$

avec:

$$\int_0^1 A_i \cos 2 i \pi x dx = 0$$

et:

$$\int_0^1 A_i x \cos 2 i \pi x dx = 0$$

on a:

$$\int_0^1 A_i x^m \cos 2 i \pi x dx = A_i m! \sum_{J=1}^M \frac{(-1)^{J+1}}{(m+1-2j)! (2 i \pi)^{2j}}, \text{ M partie entière de } \frac{m}{2}$$

$$\int_0^1 B_i x^m \sin 2 i \pi x dx = -B_i \frac{1}{2 i \pi} - \frac{m(m-1)}{(2 i \pi)^2} \int_0^1 B_i x^{m-2} \sin 2 i \pi x dx$$

avec:

$$\int_0^1 B_i \sin 2 i \pi x dx = 0 \quad \text{et} \quad \int_0^1 B_i x \sin 2 i \pi x dx = -\frac{B_i}{2 i \pi}$$

on a:

$$\int_0^1 B_i x^m \sin 2 i \pi x dx = B_i m! \sum_{J=0}^M \frac{(-1)^{J+1}}{(m-2j)! (2 i \pi)^{2j+1}} \quad \text{M partie entière de } \frac{m}{2}$$

d'où:

$$\frac{1}{m+1} + m! \sum_{i=1}^n (A_i \sum_{J=1}^M \frac{(-1)^{J+1}}{(m+1-2j)! (2 i \pi)^{2j}} + B_i \sum_{J=0}^M \frac{(-1)^{J+1}}{(m-2j)! (2 i \pi)^{2j+1}}) = m_m(x)$$

$m_m(x)$ étant le moment (non centré) d'ordre m de x .

On se trouve devant un système d'équations linéaires à $2n$ inconnues. Il est possible de le compléter en prenant $f(0) = f(1) = 0$ si l'on est sûr que la fonction de densité de x est « en cloche » et ces conditions s'écrivent :

$$\sum A_i = -1$$

6.3. Il est possible d'essayer quelque chose de plus simple, par exemple :

$$f(x) = 1 + \sum_{i=1}^n A_i \cos i \pi x$$

$$F(x) = x + \frac{1}{\pi} \sum_{i=1}^n \frac{A_i}{i} \sin i \pi x$$

(où si l'on ne conserve que les termes pairs en i , la fonction de densité devient symétrique par rapport à $x = 1/2$).

Le système d'équations linéaires permettant de calculer les coefficients A_i s'écrit

(M étant la partie entière de $\frac{m}{2}$)

$$\frac{1}{m+1} + \sum_{i=1}^n (-1)^i A_i m! \left(\sum_{j=1}^M \frac{(-1)^{j+1}}{(m+1-2j)! (i\pi)^{2j}} + (-1)^{M+1} \frac{[1-(-1)^j][1-(-1)^m]}{2 (i\pi)^{m+1}} \right) = m_m(x)$$

$m_m(x)$ étant le moment (non centré) d'ordre m de x .

6.4. En fin de compte il ne semble pas possible d'appliquer l'idée de FISHER à des développements limités du type de Fourier, et nous n'avons pas cherché à écrire de programme pour les utiliser.

7. Utilisation sur un échantillon

Les résultats obtenus avec un développement limité du type Fisher sur une loi gamma incomplète de paramètre de forme égal à 12 (coefficient d'asymétrie 0,578, d'aplatissement 0,500, assez différents des coefficients correspondant de la loi normale) nous ont fait penser que ce développement limité pourrait être utilisé avec les cumulants calculés sur un échantillon à condition d'être sûr :

- que la répartition de cet échantillon soit unimodale,
- que le rapport de la valeur du cumulants d'ordre r à la quantité $(r! K_2^{r/2})$ (K_2 étant la variance), tende rapidement vers zéro lorsque r augmente.

Ces conditions sont pratiquement toujours remplies si l'on considère la répartition d'un module interannuel ou d'une pluviométrie moyenne interannuelle par exemple (car la taille de l'échantillon servant à calculer la moyenne intervient comme prévu

par FISHER, cf. 3.3.). Elles sont souvent remplies si l'on considère la répartition de modules annuels ou de pluviométries annuelles, qui sont des moyennes ou des sommes de variables plus ou moins indépendantes: les résultats obtenus sur la loi gamma incomplète de paramètre de forme égal à 12 peuvent être considérés comme des résultats obtenus sur la somme (ou la moyenne) de 12 variables indépendantes tirées chacune d'une loi gamma incomplète de paramètre de forme égal à 1; c'est-à-dire n'ayant pas de mode réel.

Nous n'avons pas encore fait d'essais. Il est facile de les envisager compte tenu des sous-programmes mis au point (cf. 9.2.).

Annexe 1

8.1.

Polynomes orthogonaux de variables continues.

Soit a, b l'intervalle de variation de la variable x et $\theta(x)$ une « fonction de base » intégrale et positive dans le même intervalle. La suite des polynomes orthogonaux par rapport à la fonction de base et à l'intervalle a, b est la suite des polynomes $Q_n(x)$ de degré n par rapport à

$$x = Q_0 \text{ (constante)} \qquad Q_1(x) \dots \qquad Q_n(x) \dots$$

$$\text{tels que} \quad \int_a^b \theta(x) Q_n(x) Q_m(x) dx = 0 \quad \text{si } m \neq n \\ \neq 0 \text{ et positive si } m = n$$

Remarquons que la fonction de base correspond, à une constante multiplicative près, à une fonction de densité.

8.2. On voit que la définition ci-dessus ne permet d'établir chaque polynome qu'à une constante multiplicative près.

Les polynomes sont orthonormés si $\int_a^b \theta(x) Q_n^2(x) dx = 1$ quel que soit n . Cette

condition permet le calcul de la valeur absolue de la constante multiplicative, dont

$$\text{le signe s'obtient par la condition} \quad \int_a^b x \theta(x) Q_{n-1}(x) Q_n(x) dx > 0.$$

En fait nous n'utiliserons pas de polynomes orthonormés mais des polynomes orthogonaux obéissant à la condition d'avoir +1 comme coefficient du terme de plus fort degré en x .

8.3. Comme tout polynome de degré n est représentable par une somme de polynomes orthogonaux, appartenant à une même suite arbitraire, de degrés $0 \dots n$, on en déduit la méthode d'établissement de ces polynomes:

G_0 est une constante définie par $Q_0^2 \int_a^b \theta(x) dx = A_0^2$

G_1 est linéaire en x $Q_1(x) = a_1^0 Q_0 + a_1^1 x$

d'après $\int_a^b \theta(x) Q_1(x) Q_0 dx = 0$

on écrit $A_0^2 a_1^0 + a_1^1 Q_0 \int_a^b x \theta(x) dx = 0$

d'où a_1^0 si a_1^1 a été choisi.

Q_2 est du second degré en x ... etc.

8.4. Le polynôme $Q_n(x)$ admet n racines réelles et distinctes entre a et b , et entre 2 racines consécutives de $Q_n(x) = 0$
 il y a une racine et une seule de $Q_{n-1}(x) = 0$
 soit $Q_n(x)$ un polynôme orthogonal par rapport à la fonction de base $\theta(x)$ et à l'intervalle a, b et dont le coefficient de x^n est α_n .

Parmi tous les polynômes $P_n(x)$ de degré n en x , à coefficients réels, et dont le coefficient de x^n est α_n , celui pour lequel la somme $\int_a^b P_n^2(x) \theta(x) dx$ est minimale est précisément le polynôme $Q_n(x)$.

8.5. Nous donnons ci-après quelques indications sur les polynômes orthogonaux que nous avons retenus:
 intervalle de variation de la variable:
 — $\infty + \infty$: polynômes d'Hermite
 0 + ∞ : polynômes de Laguerre généralisés
 0 1 : polynômes de Jacobi

8.6. Polynômes orthogonaux d'Hermite $H_n(x)$

8.6.1. INTERVALLE DE VARIATION DE x : $-\infty, +\infty$
 Fonction de base

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

c'est-à-dire fonction de densité de la loi normale (les polynômes d'Hermite sont habituellement définis avec $e^{-\frac{x^2}{2}}$ ou e^{-x^2} comme fonction de base):

$$\int_{-\infty}^{+\infty} H_n(x) H_m(x) \theta(x) dx = 0 \quad \text{si } m \neq n$$

$$= n! \quad \text{si } m = n$$

8.6.2. EXPRESSION DE $H_n(x)$

$$H_n(x) = n! \sum_{i=0}^n (-1)^i \frac{x^{n-2i}}{i! (n-2i)! 2^i} \quad \begin{matrix} n \\ \text{N valeur entière de } - \\ 2 \end{matrix}$$

Le coefficient de x^n est 1

Les polynomes d'ordre pair sont pairs de terme constant:

$$\frac{(-1)^{n/2} n!}{(\frac{n}{2})! 2^{n/2}}$$

Les polynomes d'ordre impair sont impairs, sans termes constants.

$$H_0 = 1$$

$$H_1(x) = x$$

$$H_2(x) = x^2 - 1 \dots$$

et les polynomes suivants se calculent aisément par la relation de récurrence.

$$H_{n+1}(x) = x H_n(x) - n H_{n-1}(x)$$

8.6.3. AUTRES RELATIONS IMPORTANTES:

$$H_n''(x) - x H_n'(x) + n H_n(x) = 0$$

$$H_n'(x) - n H_{n-1}(x) = 0$$

$$\int_{-\infty}^x H_n(x) \theta(x) dx = -\theta(x) H_{n-1}(x)$$

$$\int_{-\infty}^0 H_n(x) \theta(x) dx = -\frac{2^{\frac{2n-1}{2}}}{\sqrt{\pi}} H_{n-1}(0)$$

$$\theta(x) H_n(x) = (-1)^n \frac{d^n}{dx^n} [\theta(x)]$$

8.7. Polynomes orthogonaux de LAGUERRE généralisés $L_n(x)$

8.7.1. L'intervalle de variation de x : $0, +\infty$

Fonction de base:

$$\theta(x) = \frac{1}{\Gamma(\gamma)} x^{\gamma-1} e^{-x} \quad \text{avec } \gamma > 0$$

c'est-à-dire fonction de densité de la loi gamma incomplète (les polynomes de Laguerre sont habituellement définis avec $x^\alpha e^{-x}$ comme fonction de base et avec $\frac{(-1)^n}{n!}$ comme coefficient de x^n dans $L_n(x)$).

$$\int_0^{\infty} L_n(x) L_m(x) \theta(x) dx = 0 \quad \text{si } m \neq n$$

$$= n! \frac{\Gamma(\gamma + n)}{\Gamma(\gamma)} \quad \text{si } m = n$$

8.7.2. EXPRESSION DE $L_n(x)$

$$L_n(x) = n! \sum_{i=0}^n (-1)^i \frac{\Gamma(\gamma + n) x^{n-i}}{i! (n-i)! \Gamma(\gamma + n - i)}$$

$$= n! \sum_{i=0}^n (-1)^{n-i} \frac{\Gamma(\gamma + n) x^i}{i! (n-i)! \Gamma(\gamma + i)}$$

Le coefficient de x^n est 1

Le terme constant est:

$$(-1)^n \frac{\Gamma(\gamma + n)}{\Gamma(\gamma)}$$

$$L_0 = 1 \quad L_1(x) = x - \gamma \quad L_2(x) = x^2 - 2(\gamma + 1)x + \gamma(\gamma + 1)$$

et les polynômes suivants se calculent aisément par la relation de récurrence:

$$L_{n+1}(x) = [x - (2n + \gamma)] L_n(x) - n(n + \gamma - 1) L_{n-1}(x)$$

8.7.3. AUTRES RELATIONS IMPORTANTES

$$x L_n''(x) + (\gamma - x) L_n'(x) + n L_n(x) = 0$$

$$x L_n'(x) - n L_n(x) - n(n + \gamma - 1) L_{n-1}(x) = 0$$

$$\theta(x) L_n(x) = (-1)^n \frac{d^n}{dx^n} [x^n \theta(x)] \times C^{te}$$

8.7.4. CAS PARTICULIER: POLYNOMES DE LAGUERRE, $\gamma = 1$

$$L_0 = 1 \quad L_1(x) = x - 1 \quad L_2(x) = x^2 - 4x + 2$$

$$L_{n+1}(x) = [x - (2n + 1)] L_n(x) - n^2 L_{n-1}(x)$$

8.8. Polynômes orthogonaux de Jacobi $G_n(x)$

8.8.1. INTERVALLE DE VARIATION DE $x = 0, 1$

Fonction de base:

$$\theta(x) = x^{p-1} (1-x)^{q-1} \quad \text{avec } p \text{ et } q > 0$$

La fonction de densité de la loi bêta incomplète est:

$$\frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1} \quad \text{avec} \quad \frac{1}{B(p, q)} = \frac{\Gamma(p+q)}{\Gamma(p) \Gamma(q)}$$

(les polynomes de Jacobi sont habituellement définis avec un intervalle de variation $-1, +1$ et avec $(1-x)^\alpha (1+x)^\beta$ comme fonction de base).

$$\int_0^1 G_n(x) G_m(x) \theta(x) dx = 0 \quad \text{si } m \neq n$$

$$= n! \frac{\Gamma(n+p) \Gamma(n+q) \Gamma(n+p+q-1)}{\Gamma(2n+p+q) \Gamma(2n+p+q-1)} \quad \text{si } m = n$$

8.8.2. EXPRESSION DE $G_n(x)$

$$\begin{aligned} G_n(x) &= \frac{\Gamma(n+p) n!}{\Gamma(2n+p+q-1)} \sum_{i=0}^n \frac{(-1)^i}{i! (n-i)!} \frac{\Gamma(2n+p+q-i-1)}{\Gamma(n+p+i)} x^{n-i} \\ &= \frac{\Gamma(n+p) n!}{\Gamma(2n+p+q-1)} \sum_{i=0}^n \frac{(-1)^{n-i}}{i! (n-i)!} \frac{\Gamma(n+p+q+i-1)}{\Gamma(p+i)} x^i \end{aligned}$$

Le coefficient de x^n est 1

Le terme constant est:

$$(-1)^n \frac{\Gamma(n+p)}{\Gamma(p)} \frac{\Gamma(n+p+q-1)}{\Gamma(2n+p+q-1)}$$

$G_0 = 1$ si $(p+q)$ est différent de 1

$$G_1(x) = x - \frac{p}{p+q} \quad G_2(x) = x^2 - 2 \frac{p+1}{p+q+2} x + \frac{p(p+1)}{(p+q+1)(p+q+2)}$$

et les polynomes suivants se calculent aisément par la relation de recurrence:

$$\begin{aligned} G_{n+1}(x) &= \left[x - \frac{2n(n+p+q-1) + p(p+q-2)}{(2n+p+q-2)(2n+p+q)} \right] G_n(x) \\ &\quad - \frac{n(n+p-1)(n+q-1)(n+p+q-2)}{(2n+p+q-3)(2n+p+q-2)^2(2n+p+q-1)} G_{n-1}(x) \end{aligned}$$

8.8.3. AUTRES RELATIONS IMPORTANTES

$$x(1-x) G_n''(x) + [p - (p+q)x] G_n'(x) + n(n+p+q-1) G_n(x) = 0$$

$$x(1-x) G_n'(x) + \left[nx - \frac{n(n+q-1)}{2n+p+q-2} \right] G_n(x)$$

$$- \frac{n(n+p-1)(n+q-1)(n+p+q-2)}{(2n+p+q-3)(2n+p+q-2)^2} G_{n-1}(x) = 0$$

$$\theta(x) G_n(x) = (-1)^n \frac{d^n}{dx^n} [x^n (1-x)^n \theta(x)] \times C^{te}$$

8.8.4. CAS PARTICULIER POLYNOMES DE LEGENDRE $p = q = 1$

$$G_0 = 1 \quad G_1(x) = x - \frac{1}{2} \quad G_2(x) = x^2 - x + \frac{1}{6} \quad , \quad \frac{1}{B(p, q)} = 1$$

$$G_{n+1}(x) = (x - \frac{1}{2}) G_n(x) - \frac{n^2}{4(4n^2 - 1)} G_{n-1}(x)$$

8.8.5. CAS PARTICULIER POLYNOMES DE TCHEBYSCHOFF DE PREMIÈRE ESPÈCE $p=q=1/2$

$$G_0 = 2 \quad G_1(x) = x - \frac{1}{2} \quad G_2(x) = x^2 - x + \frac{1}{8} \quad , \quad \frac{1}{B(p, q)} = \frac{1}{\pi}$$

$$G_{n+1}(x) = (x - \frac{1}{2}) G_n(x) - \frac{1}{16} G_{n-1}(x)$$

les valeurs nulles des polynomes de première espèce sont obtenues pour :

$$x = \frac{1}{2} \left(1 - \cos \frac{2K-1}{2n} \pi \right) \quad K \text{ de } 1 \text{ à } n$$

Les maxima et les minima des valeurs du polynome de degré n sont atteints, dans l'intervalle $0,1$, pour les valeurs

$$x_K = \frac{1}{2} \left(1 - \cos \frac{K\pi}{n} \right) \quad K \text{ de } 1 \text{ à } (n-1) \quad \text{et sont} \quad \frac{(-1)^k}{2^{n-1}}$$

Le polynome de première espèce de degré n est, parmi tous les polynomes possibles de degré n dont le coefficient de x^n est égal à 1, celui pour lequel la valeur absolue maximale atteinte dans l'intervalle $0,1$ est la plus petite.

8.8.6. CAS PARTICULIER POLYNOMES DE TCHEBYSCHOFF DE DEUXIÈME ESPÈCE $p = q = \frac{3}{2}$

$$G_0 = 1 \quad G_1(x) = x - \frac{1}{2} \quad G_2(x) = x^2 - x + \frac{3}{16} \quad , \quad \frac{1}{B(p, q)} = \frac{8}{\pi}$$

$$G_{n+1}(x) = \left(x - \frac{1}{2} \right) G_n(x) - \frac{3}{16} G_{n-1}(x)$$

Les valeurs nulles des polynomes de deuxième espèce sont obtenues pour :

$$x = \frac{1}{2} \left(1 - \cos \frac{K}{n+1} \pi \right) \quad K \text{ de } 1 \text{ à } n$$

Annexe 2

9. Moments et cumulants

9.1.

Nous supposons, ci-dessous, que les moments ou les cumulants de la variable x sont connus à priori, indépendamment d'un échantillon quelconque de taille finie.

Le moment (centré ou non) d'ordre l est noté m_l

Le cumulants d'ordre l est noté K_l

9.1.1. CALCUL DES MOMENTS A PARTIR DES CUMULANTS

le moment

$$m_l = l! \sum \frac{1}{\pi j!} \Pi \left(\frac{K_i}{i!} \right)^j$$

La valeur maximale de i est l

La valeur minimale de i est 1 si $K_1 \neq 0$ et l'on calcule les moments non centrés. Si l'on prend $K_1 = 0$, la valeur minimale de i est 2 et l'on calcule les moments centrés. Dans chaque terme de la somme $\sum i j = l$, avec $0 \leq j \leq l$

$$m_1 = K_1$$

$$m_2 = K_2 + K_1^2$$

$$m_3 = K_3 + 3 K_2 K_1 + K_1^3$$

$$m_4 = K_4 + 4 K_3 K_1 + 3 K_2^2 + 6 K_2 K_1^2 + K_1^4 \text{ etc.}$$

Le sous-programme MPARK calcule les moments d'après les cumulants, jusqu'à l'ordre 56 au plus car il utilise la fonction gamma. Il est très possible d'obtenir un dépassement de la capacité d'un ordinateur IBM 360 avant même d'atteindre l'ordre 56.

9.1.2. CALCUL DES CUMULANTS A PARTIR DES MOMENTS

Le cumulants

$$K_l = l! \sum \frac{(-1)^m m!}{\pi j!} \Pi \left(\frac{m_i}{i!} \right)^j$$

La valeur maximale de i est l

La valeur minimale de i est 1 si les moments ne sont pas centrés, ou 2 si les moments sont centrés ($m_1 = 0$)

dans chaque terme de la somme $\sum i j = l$ avec $0 \leq j \leq l$

et

$$m = \sum j - 1$$

$$K_1 = m_1$$

$$K_2 = m_2 - m_1^2$$

$$K_3 = m_3 - 3 m_2 m_1 + 2 m_1^3$$

$$K_4 = m_4 - 4 m_3 m_1 - 3 m_2^2 + 12 m_2 m_1^2 - 6 m_1^4 \text{ etc.}$$

Si l'on part des moments centrés, il faut donner à K_1 la valeur moyenne de la variable x

Le sous-programme KPARM calcule les cumulants d'après les moments, jusqu'à l'ordre 56 au plus (même remarque que pour le sous-programme MPARK).

9.1.3. CALCUL DES MOMENTS AVEC CHANGEMENT D'ORIGINE

Le moment d'ordre l de la variable réduite $u = \frac{x - x_0}{s}$ (s étant un paramètre d'échelle, et x_0 un paramètre de position: moyenne, borne...) se calcule d'après les moments d'ordre l et inférieurs de la variable x par

$$m_{nl} = \frac{1}{s^l} \left[m_{xl} + (-1)^l x_0^l + \sum_{i=1}^{l-1} (-1)^i \frac{l!}{i! (l-i)!} m_{x(l-i)} x_0^i \right]$$

En fortran on pourrait l'écrire:

$$\text{CMU}(L) = \text{CMX}(L)$$

$$A = 1.$$

$$DØ 1 IL = 2, L$$

$$A = -A * (L + 2 - IL) / (IL - 1) * X Ø$$

$$1 \text{ CMU}(L) = \text{CMU}(L) + A * \text{CMX}(L + 1 - IL)$$

$$\text{CMU}(L) = (\text{CMU}(L) + X Ø ** L * (1 - 2 * (MØD(L, 2)))) / S ** L$$

9.2. Les paragraphes 9.1.1. et 9.1.2. donnent des relations qui, répétons-le, ne sont utilisables que dans les cas où l'on connaît à priori tous les cumulants ou tous les moments de la distribution. Nous allons maintenant définir le cumul des moments et des cumulants d'après un échantillon de taille n de valeur connues de la variable x .

9.2.1. CALCUL DES MOMENTS

Le moment d'ordre l déduit d'un échantillon de taille n est, quelque soit l ,

$$m_{xl} = \frac{1}{n} \sum_{i=1}^n x_i^l$$

et dans le cas où ni x_0 ni s ne sont calculés d'après l'échantillon, la relation du paragraphe 9.1.3. peut être utilisée pour calculer les moments de la variable $u = \frac{x - x_0}{s}$

9.2.2. CALCUL DES CUMULANTS

Le cumulatif d'ordre l déduit d'un échantillon de taille n est, pour $l \leq n$

$$K_l = l! \sum_{j=1}^l \frac{(-1)^{j-1} (j-1)!}{n(n-1)\dots(n-j+1)} \sum \frac{x_1^{p_1} x_2^{p_2} \dots x_s^{p_s}}{(p_1!)^{\pi_1} \dots (p_s!)^{\pi_s} \pi_1! \dots \pi_s!}$$

dans la seconde somme J est le nombre de facteurs, p_m est la puissance à laquelle x_1 est élevée, π_m est le nombre de fois où la puissance p_m est répétée, et $J = \sum \pi_m$ avec $l = \sum p_m \pi_m$.

Dans chaque terme de cette seconde somme les indices des x ne doivent pas être répétés. Cette seconde somme est étendue à toutes les combinaisons possibles des p_m et des π_m conservant la valeur de J (et celle de l)

ainsi pour $l = 6$ et $j = 3$ par exemple on aura 3 types de produits:

p valant 1 1 4, π valant 2 1
 p valant 1 2 3, π valant 1 1 1
 p valant 2 2 2, π valant 3

dans le premier type de produit on aura les termes:

$$\begin{array}{l} x_1 x_2 x_3^4 \\ x_2 x_1 x_3^4 \\ x_3 x_2 x_1^4 \\ x_2 x_3 x_1^4 \dots \end{array}$$

La sous-routine KPARX effectue ces calculs en évitant les répétitions du genre des deux premiers termes écrits ci-dessus, ce qui supprime les divisions par les $\pi_m!$. Elle pourrait fonctionner jusqu'à l'ordre 56 (même remarque que pour la sous-routine MPRAK), mais le temps de travail serait prohibitif (cf. nota).

9.2.3.

Les 6 premiers cumulants peuvent s'écrire, en fonction des sommes S_l des puissances l des x_i : $S_l = \sum x_i^l = n \cdot m_l$

$$K_1 = \frac{1}{n} S_1$$

$$K_2 = \frac{1}{n(n-1)} [n S_2 - S_1^2]$$

$$K_3 = \frac{1}{n(n-1)(n-2)} [n(n S_3 - 3 S_2 S_1) + 2 S_1^3]$$

$$K_4 = \frac{1}{n(n-1)(n-2)(n-3)} [n(n+1)(n S_4 - 4 S_3 S_1) - 3 n(n-1) S_2^2 + 6 S_1^2 (2n S_2 - S_1^2)]$$

Nota. — Le nombre de calculs élémentaires effectués par la sous-routine KPARX reste impressionnant. Il lui faut un peu plus d'une minute sur un ordinateur IBM 360-75 pour calculer les cinq premiers cumulants déduits d'un échantillon de 13 valeurs seulement.

$$K_5 = \frac{1}{n(n-1)(n-2)(n-3)(n-4)} [n^2(n+5)(nS_5 - 5S_4S_1) - 10n^2(n-1)S_3S_2 \\ + 20n(n+2)S_3S_1^2 + 30n(n-1)S_2^2S_1 \\ - 12S_1^3(5nS_2 - 2S_1^2)]$$

$$K_6 = \frac{1}{n(n-1)(n-2)(n-3)(n-4)(n-5)} [n(n+1)n^2 + 15n-4)(nS_6 - 6S_5S_1) \\ - 15n(n-1)^2(n+4)S_4S_2 \\ - 10n(n-1)(n^2-n+4)S_3^2 \\ + 30n(n^2+9n+2)S_4S_1^2 \\ + 120n(n^2-1)S_3S_2S_1 \\ + 30n(n-1)(n-2)S_2^3 - 120n(n+3)S_3S_1^3 \\ - 270n(n-1)S_2^2S_1^2 + 120S_1^4(3nS_2 - S_1^2)]$$

La sous-routine KPARSX calcule, d'après ces formules les six premiers cumulants en partant des n valeurs x_i de l'échantillon.

9.2.4.

Les cumulants ayant été calculés d'après les valeurs de l'échantillon, on peut alors calculer les moments centrés (sur la valeur moyenne déduite de l'échantillon) par les relations du paragraphe 9.1.1.

10. Emploi des développements limités

Nous n'avons pas eu la prétention d'épuiser toutes les possibilités de développement limités de fonctions de distributions, ni de passer en revue tous les polynomes orthogonaux.

Etant donné les remarques faites aux paragraphes 4.6. et 5.7., nous n'avons écrit de programmes que pour les développements limités du type de Fisher (cf. § 3.3.) qui nous semblent utilisables dans les cas suivants:

10.1.

La fonction de répartition est connue par l'ensemble de ses moments ou de ses cumulants, sans qu'on puisse écrire son expression mathématique: exemple: répartition de la moyenne d'un échantillon de taille connue n , tiré d'une population mère connue, distribuée par exemple suivant une loi exponentielle généralisée, une loi gaussio-logarithmique, une loi de Gumbel...

Dans les deux premiers cas les moments de la population mère sont d'expressions simples, on peut en déduire les cumulants par le sous-programme KPARM. Dans le troisième cas les cumulants sont directement d'expressions simples. De toute façon le cumulants d'ordre r de la distribution de la moyenne est égal au cumulants d'ordre r de la distribution de la population mère divisé par n^{r-1} .

Le sous-programme SPARKF calcule les coefficients à rentrer dans le sous-programme FDEVS (de calcul de la probabilité au non dépassement). Il semble inutile de dépasser, même si n n'est pas grand, 30 coefficients ($L = 30$ dans le sous-programme SPARKF, $M = 30$ dans le sous-programme FDEVS), ce qui utilise les douze premiers cumulants seulement ($LR = 12$ dans le sous-programme KPARM)

Le sous-programme VDEVS permet d'inverser la fonction de distribution (avec la même valeur de M que pour FDEVS).

10.2.

Etude de la répartition de la moyenne d'un échantillon de taille n uniquement d'après les valeurs de l'échantillon, sans aucune hypothèse sur la loi de répartition de la population mère (cf. 7.). La sous-routine KPARSX calcule les six premiers cumulants d'après l'échantillon de taille n. Le cumulatif d'ordre r de la moyenne est égal au cumulatif d'ordre r calculé par KPARSX, divisé par n^{r-1}.

Le sous-programme SPARKF calcule (L = 12) d'après les cumulants de la moyenne, les coefficients à rentrer dans le sous-programme FDEVS (où M = 12).

Il est possible d'améliorer le sous-programme KPARSX pour lui faire calculer les huit premiers cumulants. Les valeurs de L et de M deviendraient égales à 18.

```

SUBROUTINE KPARSX(C,LE)
CALCUL DES MOMENTS CENTRES JUSQU' A L'ORDRE L APRES LES CUMULANTS
CUMULANTS P, JUSQU' A L'ORDRE L. LES MOMENTS SONT DETRUITES
SI L'ORDRE DES ELEMENTS CALCULES SONT LES MOMENTS CENTRES.
LE PROGRAMME PRINCIPAL FOURNIT LE ET LES C EN DOUBLE PRECISION
LE SOUS-PROGRAMME DOIT ETRE DIMENSIONNE
AU MOINS LX POUR KK ET Y EN DOUBLE PRECISION POUR P ET S.
DOUBLE PRECISION C(1P),P(1P),SS,SEF
DIMENSION X(1P)
C(1)=C(1)
SS=1.00
DO 10 I=2,L
  SS=SS*(1+1/I)
10 C(I)=C(I)/SS
DO 24 IR=2,LR
  C(1P)=C(1P)
  DO 22 IR=2,LR
    SS=0.
    DO 11 I=1,IR
      SS=SS*(1+1/I)
11 X(I)=C(1)
DO 13 I=2,IR
  SS=SS*(1+1/I)
13 X(I)=X(I)+1/I
14 X(IR)=X(IR)+1/I
DO 12 IR=2,LR
  SS=SS*(1+1/I)
  F=1.00
  DO 17 IR=2,IR
    F=(F*(1+1/I)+X(IR)-1)/F
17 SS=SS*(F*(1+1/I))
  SS=SS*SS
19 I=1
  IF(I.EQ.0)GO TO 22
  IF(X(1).EQ.XX(1+1))GO TO 17
  DO 20 IR=1,IR
    X(IR)=X(IR)+X(1)
  GO TO 17
22 C(1P)=C(1P)+X(1)
24 C(1P)=C(1P)+X(1)
RETURN
END

SUBROUTINE SPARKF(K,L)
CALCUL EN DOUBLE PRECISION DES CUMULANTS P JUSQU' A L'ORDRE L
APRES UN ECHANTILLON DE TAILLE N DE LA VARIABLE X FOURNI PAR
LE PROGRAMME PRINCIPAL FOURNIT LE ET LES C EN DOUBLE PRECISION
DIMENSION X(1P)
DO 10 I=2,L
  SS=1.00
  DO 11 I=1,I-1
    SS=SS*(1+1/I)
10 C(I)=C(I)/SS
DO 24 IR=2,LR
  C(1P)=C(1P)
  DO 22 IR=2,LR
    SS=0.
    DO 11 I=1,IR
      SS=SS*(1+1/I)
11 X(I)=C(1)
DO 13 I=2,IR
  SS=SS*(1+1/I)
13 X(I)=X(I)+1/I
14 X(IR)=X(IR)+1/I
DO 12 IR=2,LR
  SS=SS*(1+1/I)
  F=1.00
  DO 17 IR=2,IR
    F=(F*(1+1/I)+X(IR)-1)/F
17 SS=SS*(F*(1+1/I))
  SS=SS*SS
19 I=1
  IF(I.EQ.0)GO TO 22
  IF(X(1).EQ.XX(1+1))GO TO 17
  DO 20 IR=1,IR
    X(IR)=X(IR)+X(1)
  GO TO 17
22 C(1P)=C(1P)+X(1)
24 C(1P)=C(1P)+X(1)
RETURN
END

```

