

Systeme de Pearson

Description mathématique du système de Pearson
pour son implémentation en C++
à l'aide de la librairie GSL

Stéphane Derrode

École Centrale Marseille (ECM),
Groupe Signaux Multidimensionnels, Institut Fresnel (CNRS - UMR 6133),
Domaine universitaire de Saint Jérôme, 13013 Marseille Cedex 20, France
int+ 33 4 91 05 47 30, stephane.derrode@centrale-marseille.fr

9 décembre 2009

Table des matières

1 Synthèse sur le système de Pearson	3
1.1 Présentation du système de Pearson	3
1.2 Caractérisation du système de Pearson	3
1.2.1 Notations : moments, skewness et kurtosis	3
1.2.2 Caractérisation des paramètres du système	4
1.3 Description du diagramme de Pearson	5
1.3.1 Limites inférieure et supérieure	5
1.3.2 Courbes f_{III} et f_{XII}	5
1.3.3 Courbes f_V	5
1.3.4 Courbes f_I	6
1.3.5 Quelques points particuliers	7
1.4 Expressions des densités de probabilités	7
1.4.1 Type XIII (lois Gaussiennes)	7
1.4.2 Type III (lois Gamma) et Type X (lois exponentielles)	8
1.4.3 Type I (lois Béta de 1 ^{re} espèce)	9
1.4.4 Type II (cas particulier des lois Béta de 1 ^{re} espèce)	10
1.4.5 Type VI (lois Béta de 2 ^e espèce)	11
1.4.6 Type IV	12
1.4.7 Type V (lois Gamma inverse)	13
1.4.8 Type VII (lois de Student't)	14
1.5 Illustrations des formes dans le système de Pearson	15
2 Annexes	15
2.1 Expressions des lois usuelles	15
2.2 Quelques fonctions spéciales	16

1 Synthèse sur le système de Pearson

La plupart des développements présentés dans cette section sont issus du livre de Johnson et Kotz [1, 2], du livre de Miller [3], et du livre de Stuart et Ord [4]. Je me suis également servi des travaux de Delignon et Marzouki [5, 6] et du site web de Eric Weisstien¹.

1.1 Présentation du système de Pearson

Une distribution f de \mathbb{R} appartient au système de Pearson si elle satisfait l'équation différentielle suivante² :

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{d \ln f(x)}{dx} = -\frac{x+a}{c_0 + c_1x + c_2x^2} \quad (1)$$

La forme générale de la solution de l'équation (1) s'écrit (site Web de Eric Weisstein) :

$$f(x) = C \frac{1}{(c_0 + c_1x + c_2x^2)^{\frac{1}{2c_2}}} \exp \left[\frac{(c_1 - 2ac_2) \arctan \left(\frac{c_1 + 2c_2x}{\sqrt{4c_0c_2 - c_1^2}} \right)}{c_2 \sqrt{4c_0c_2 - c_1^2}} \right] \quad (2)$$

L'existence de la solution dépend :

- des racines du polynôme de $c_0 + c_1x + c_2x^2$, de déterminant $\delta = c_1^2 - 4c_0c_2$, et
- de la valeur de c_2 (positive, négative ou nulle),

La résolution de cette équation différentielle amène à considérer 8 familles de lois solutions, formant ce que l'on appelle le « système de Pearson »³.

1.2 Caractérisation du système de Pearson

L'intérêt du système de Pearson est qu'il couvre une large gamme de formes des distributions avec un nombre très limité de paramètres puisque les 4 premiers moments ou bien le skewness et le kurtosis suffisent.

1.2.1 Notations : moments, skewness et kurtosis

Soit Y une variable aléatoire réelle dont la loi appartient au système de Pearson. Nous noterons :

- μ_q les moments centrés de Y ($\mu_1 = 0$), d'ordre q inférieur à 4 (en particulier, nous noterons $\sigma = \sqrt{\mu_2}$),
- γ_1 le « skewness » et γ_2 le « kurtosis » de Y selon la définition formulée par Fisher :

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}, \quad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3. \quad (3)$$

Des définitions sensiblement différentes du « skewness » et du « kurtosis » ont été proposées par Pearson :

$$\beta_1 = \gamma_1^2, \quad \beta_2 = \gamma_2 + 3. \quad (4)$$

Nous serons amenés à jongler avec les deux types de définitions pour simplifier les expressions.

¹— Eric Weisstein's world of Mathematics : <http://mathworld.wolfram.com/PearsonSystem.html>, a Wolfram web resource.

²— *Remarque* : cette équation a typiquement la forme d'une intégrale de Mellin, cf. travaux de Henri Nicolas et al, ENST Paris. Ils utilisent les tables d'intégrales de Mellin données dans le livre de Colombo [7]. À voir ...

³— À l'origine, Pearson a extrait 13 familles solutions, mais il est apparu que certaines d'entre-elles étaient des cas particuliers des précédentes (forme en « U », en « L » ou en « J »). Ainsi, 8 familles suffisent à décrire le système.

1.2.2 Caractérisation des paramètres du système

Les paramètres a , c_0 , c_1 et c_2 de l'équation différentielle (1) peuvent s'exprimer en fonction des moments des distributions de la manière suivante⁴ :

$$\begin{cases} c_0 = \mu_2 \frac{4\beta_2 - 3\beta_1}{X} = \mu_2 \frac{4\mu_2\mu_4 - 3\mu_3^2}{10\mu_2\mu_4 - 12\mu_3^2 - 18\mu_2^3} \\ c_1 = \sigma \frac{\gamma_1 (\beta_2 + 3)}{X} = a = \mu_3 \frac{\mu_4 + 3\mu_2^2}{10\mu_2\mu_4 - 12\mu_3^2 - 18\mu_2^3} \\ c_2 = \frac{2\beta_2 - 3\beta_1 - 6}{X} = \frac{2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3}{10\mu_2\mu_4 - 12\mu_3^2 - 18\mu_2^3} \end{cases} \quad (5)$$

avec $X = 10\beta_2 - 12\beta_1 - 18$. X est nul pour $\beta_2 = \frac{3}{5} (2\beta_1 + 3)$, ce qui correspond à la loi de type XII, que nous évoquons dans la section 1.4.3.

Nous serons souvent amenés par la suite à considérer le coefficient Δ défini de la manière suivante :

$$\Delta = \frac{2\beta_2 - 3\beta_1 - 6}{\beta_2 + 3} = \sigma\gamma_1 \frac{c_2}{c_1} \quad (6)$$

En remarquant que :

$$\Delta + 2 = \frac{4\beta_2 - 3\beta_1}{\beta_2 + 3} = \frac{\gamma_1}{\sigma} \frac{c_0}{c_1}, \quad \frac{\Delta}{\Delta + 2} = \frac{2\beta_2 - 3\beta_1 - 6}{4\beta_2 - 3\beta_1} = \mu_2 \frac{c_2}{c_0},$$

X s'écrit alors $X = 2(\beta_2 + 3) (1 + 2\Delta)^5$. Nous pouvons ré-écrire (5) de la manière suivante :

$$\begin{cases} c_0 = \mu_2 \frac{\Delta + 2}{2(1 + 2\Delta)} = \mu_2 (1 - 3c_2) \\ c_1 = \sigma \frac{\gamma_1}{2(1 + 2\Delta)} = a \\ c_2 = \frac{\Delta}{2(1 + 2\Delta)} \end{cases} \quad (7)$$

On peut donc ré-écrire δ sous la forme suivante :

$$\delta = \mu_2 \frac{\beta_1 - 4(\Delta + 2)\Delta}{4(1 + 2\Delta)^2}.$$

En notant :

$$\kappa = \frac{\beta_1 (\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} = \frac{\beta_1}{4\Delta(\Delta + 2)} = \frac{c_1^2}{4c_0c_2}, \quad (8)$$

nous obtenons les solutions suivantes de l'équation différentielle (1) :

Type	Conditions	Famille de lois
Type I (et XII)	$\kappa < 0$	Distribution bêta
Type II	$\beta_1 = 0, \beta_2 < 3$	Type I avec racines symétriques
Type III (X)	$2\beta_2 - 3\beta_1 - 6 = 0$	Distribution gamma
Type IV	$0 < \kappa < 1$	Sans nom particulier
Type V	$\kappa = 1$	Distribution Gamma inverse
Type VI	$\kappa > 1$	Distribution bêta inverse
Type VII	$\beta_1 = 0, \beta_2 > 3$	Distribution Students't (Type VII)
Type XIII	$\beta_1 = 0, \beta_2 = 3$	Distribution gaussienne

⁴— Le fait que $a = c_1$ provient du fait que le système considéré est centré ($\mu_1 = 0$).

⁵— Le cas $X = 0$ que nous avons évoqué ci-dessus correspond à $1 + 2\Delta = 0$, soit $\Delta = -\frac{1}{2}$.

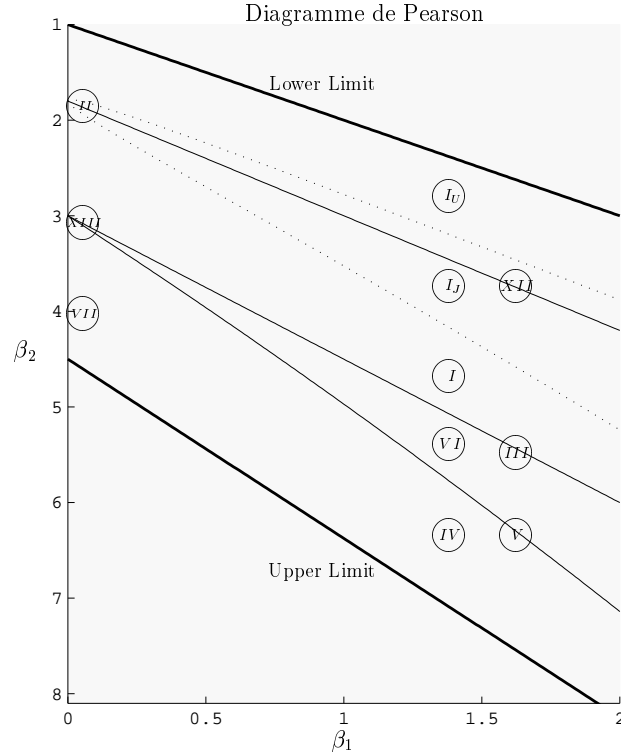


FIG. 1 – Caractérisation des lois du système de Pearson dans le graphe (β_1, β_2) . La distinction entre les familles de lois I_U , I_J et I évoque les différentes formes caractéristiques des lois bêta.

Les huit familles de lois solutions de l'équation différentielle sont classiquement représentées dans le « Diagramme de Pearson » (cf. figure 1), en fonction des valeurs de β_1 et β_2 . Les expressions de toutes les lois sont données dans la section 1.4. Voyons auparavant les limites du diagramme.

1.3 Description du diagramme de Pearson

La figure 2 représente le diagramme de Pearson pour de grandes valeurs de β_1 et β_2 (le cadre en pointillé rappelle l'échelle de la figure 1). L'expression analytique des courbes de séparation entre les différentes familles de lois sont les suivantes :

1.3.1 Limites inférieure et supérieure

$$\begin{aligned} f_{Low}(\beta_2) &= \beta_1 + 1. \\ f_{Up}(\beta_2) &= \frac{15}{8}\beta_1 + \frac{9}{2}. \end{aligned}$$

1.3.2 Courbes f_{III} et f_{XII}

$$\begin{aligned} f_{III}(\beta_2) &= \frac{3}{2}(\beta_1 + 2). \\ f_{XII}(\beta_2) &= \frac{3}{5}(2\beta_1 + 3). \end{aligned}$$

1.3.3 Courbes f_V

On obtient les courbes f_V en résolvant l'équation :

$$\beta_1 (\beta_2 + 3)^2 = 4 (4\beta_2 - 3\beta_1) (2\beta_2 - 3\beta_1 - 6).$$

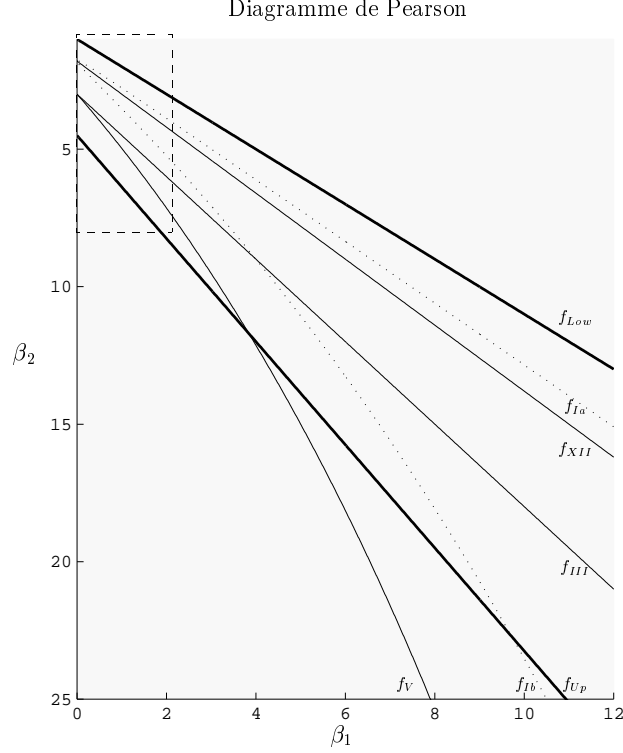


FIG. 2 – Vue d'ensemble du diagramme de Pearson - Courbes de séparation des familles de lois.

Les solutions sont données par les deux courbes à valeurs réelles f_{V_a} et f_{V_b} :

$$\begin{cases} f_{V_a}(\beta_2) = \frac{3}{\beta_1 - 32} \left(-13\beta_1 - 16 - 2(\beta_1 + 4)^{\frac{3}{2}} \right) \\ f_{V_b}(\beta_2) = \frac{3}{\beta_1 - 32} \left(-13\beta_1 - 16 + 2(\beta_1 + 4)^{\frac{3}{2}} \right) \end{cases} \quad (9)$$

pour $\beta_1 \neq 32$.

La courbe f_{V_a} est à l'intérieur des limites du système pour $\beta_1 \leq \frac{96}{25}$ ($\beta_2 = \frac{117}{10}$). La courbe f_{V_b} est, quand à elle, constamment à l'extérieur du domaine.

1.3.4 Courbes f_I

On obtient les courbes f_I en résolvant l'équation :

$$4(4\beta_2 - 3\beta_1)(5\beta_2 - 6\beta_1 - 9)^2 = \beta_1(\beta_2 + 3)^2(8\beta_2 - 9\beta_1 - 12).$$

Les solutions sont données par les trois courbes f_{I_a} , f_{I_b} et f_{I_c} :

$$\begin{cases} f_{I_a}(\beta_2) = \frac{3}{16} \left(-\frac{B^{\frac{1}{3}}}{\beta_1 - 50} - C + 2D + I\sqrt{3} \left(\frac{B^{\frac{1}{3}}}{\beta_1 - 50} - C \right) \right) \\ f_{I_b}(\beta_2) = \frac{3}{16} \left(-\frac{B^{\frac{1}{3}}}{\beta_1 - 50} - C + 2D - I\sqrt{3} \left(\frac{B^{\frac{1}{3}}}{\beta_1 - 50} - C \right) \right) \\ f_{I_c}(\beta_2) = \frac{3}{8} \left(\frac{B^{\frac{1}{3}}}{\beta_1 - 50} + C + D \right) \end{cases} \quad (10)$$

en notant :

$$\begin{cases} A = \beta_1(\beta_1 + 4)^8 \\ B = (\beta_1^2 - 360\beta_1 + 2000)(\beta_1 + 4)^4 + 32I\sqrt{A}(\beta_1 - 50) \\ C = \frac{(\beta_1 + 100)(\beta_1 + 4)^3}{(\beta_1 - 50)B^{\frac{1}{3}}} \\ D = \frac{\beta_1^2 - 144\beta_1 - 160}{\beta_1 - 50} \end{cases} \quad (11)$$

Ces trois courbes sont à valeurs réelles (si, si!!!).

1.3.5 Quelques points particuliers

- La courbe f_{I_b} coupe f_{III} au point de coordonnées $(\beta_1 = 4, \beta_2 = 9)$.
- Les courbes f_{I_a} et f_{I_b} ont pour ordonnée à l'origine $\beta_2 = \frac{9}{5}$.

Voyons maintenant la description précise (i.e. en fonction de a, c_0, c_1 et c_2) des huit familles intervenant dans le système de Pearson.

1.4 Expressions des densités de probabilités

Nous donnons ici l'expression des densités de probabilités et des fonctions de répartition lorsqu'elles sont connues, des huit familles de lois du système de Pearson (notées de p_1 à p_8). Chaque famille de lois sera illustrée par quelques figures représentant les formes caractéristiques que l'on peut rencontrer. Tous les exemples seront présentés avec $\gamma_1 > 0$. La dernière sous-section illustrant le cas $\gamma_1 < 0$. Lorsqu'il est possible, nous avons également écrit comment générer des nombres aléatoires suivants ces distributions, à partir des fonctions de la librairie GSL⁶. Il s'agit en général de simple changement de variables.

Les expressions sont entièrement déterminées par les coefficients :

- $a = c_1, c_0$ et c_2 , ou
- μ_2, μ_3 et μ_4 , ou bien encore
- μ_2, β_1 et β_2 .

Le tableau 2 rappelle les expressions des densités de probabilité (centrées, notées f^1 à f^4) de certaines lois usuelles, tandis que le tableau 3 donne les expressions des fonctions spéciales intervenant dans les calculs.

1.4.1 Type XIII (lois Gaussiennes)

Dans ce cas, nous avons $\beta_1 = 0$ et $\beta_2 = 3$, soit $\mu_1 = 0, \mu_2 = \sigma^2, \mu_3 = 0$ et $\mu_4 = 3\sigma^4$.

Nous pouvons alors écrire ($X = 12, \Delta = 0$) :

$$\begin{cases} c_0 = \mu_2 \\ c_1 = a = c_2 = 0 \end{cases}$$

et l'équation différentielle s'écrit :

$$\frac{d \ln p_{13}(x)}{dx} = -\frac{x}{c_0}$$

La solution s'écrit finalement :

$$p_{13}(x) = \frac{1}{\sqrt{2\pi c_0}} e^{\left(-\frac{1}{2} \frac{x^2}{c_0}\right)} = f_{\sqrt{c_0}}^1(x)$$

⁶— La librairie GSL (« Gnu Scientific Library ») est une librairie scientifique programmée en C : <http://sources.redhat.com/gsl/>.

La fonction de répartition s'écrit alors :

$$P_{13}(x) = \frac{1}{2} \operatorname{Erfc} \left(\frac{-x}{\sqrt{2c_0}} \right)$$

La génération de nombres aléatoires est réalisée grâce à :

$$Va_{13} = \operatorname{gsl_ran_gaussian}(R, \sqrt{c_0}) + \mu_1$$

Les paramètres utilisés dans la figure 3 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Gaussienne	+40.00	+100.00	+0.00	+30000.00	+0.00	+3.00

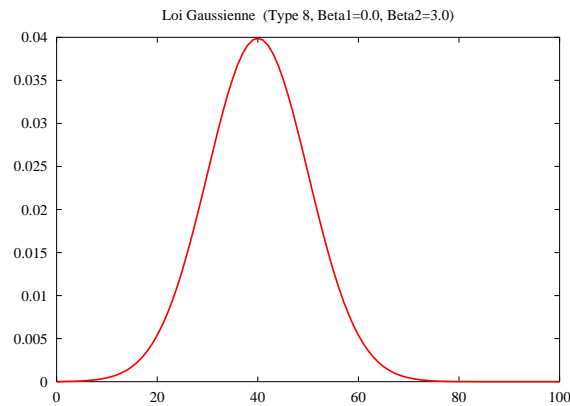


FIG. 3 – Loi gaussienne.

1.4.2 Type III (lois Gamma) et Type X (lois exponentielles)

Dans ce cas, nous avons $2\beta_2 - 3\beta_1 - 6 = 0$ et nous pouvons alors écrire ($\Delta = 0$) :

$$\begin{cases} c_0 = \mu_2 \\ c_1 = \sigma \frac{\gamma_1}{2} = \frac{\mu_3}{2\mu_2} = a \\ c_2 = 0 \end{cases}$$

et l'équation différentielle s'écrit :

$$\frac{d \ln p_3(x)}{dx} = -\frac{c_1 + x}{c_0 + c_1 x}$$

La solution de l'équation différentielle s'écrit :

$$p_3(x) = C_3 e^{\left(-\frac{x}{c_1}\right)} (c_0 + c_1 x)^{\frac{c_0}{c_1} - 1}$$

avec C_3 tel que $\int p_3(x) dx = 1$. En effectuant le changement de variable $y = c_0 + c_1 x$, nous obtenons :

$$p_3(x) = C_3 e^{\left(-\frac{y-c_0}{c_1}\right)} y^{\frac{c_0}{c_1} - 1} = C_3 e^{-\frac{y}{c_1}} y^{\frac{c_0}{c_1} - 1}$$

Soit, avec $\alpha = \frac{c_0}{c_1}$ et $\theta = c_1^2$, nous obtenons finalement :

$$p_3(x) = C_3 e^{-\frac{y}{\theta}} y^{\alpha - 1} = |c_1| f_{\alpha, \theta}^2(c_0 + c_1 x).$$

Cette équation existe pour ($\gamma_1 > 0$ et $x \geq -\frac{c_0}{c_1}$) ou pour ($\gamma_1 < 0$ et $x \leq -\frac{c_0}{c_1}$). Sous les mêmes conditions, la fonction de répartition s'écrit :

$$P_3(x) = \Gamma_\alpha \left(\frac{c_0 + c_1 x}{\theta} \right).$$

Cas particulier : Il intervient lorsque $c_0 = c_1^2$, soit $\beta_1 = 4$ et $\beta_2 = 9$, soit $\alpha = 1$. Ainsi, la loi Gamma devient une loi exponentielle (Type X selon la classification de Pearson) de paramètre $\theta = c_1^2 = c_0 = \mu_2$.

La génération de nombres aléatoires est réalisée grâce à :

$$Va_3 = \frac{1}{c_1} \text{gsl_ran_gamma}(R, \alpha, \theta) + \mu_1$$

Les paramètres utilisés dans la figure 4 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Gamma	+40.00	+100.00	+707.11	+37500.00	+0.50	+3.75
Exponentielle	+40.00	+100.00	+2000.00	+90000.00	+4.00	+9.00

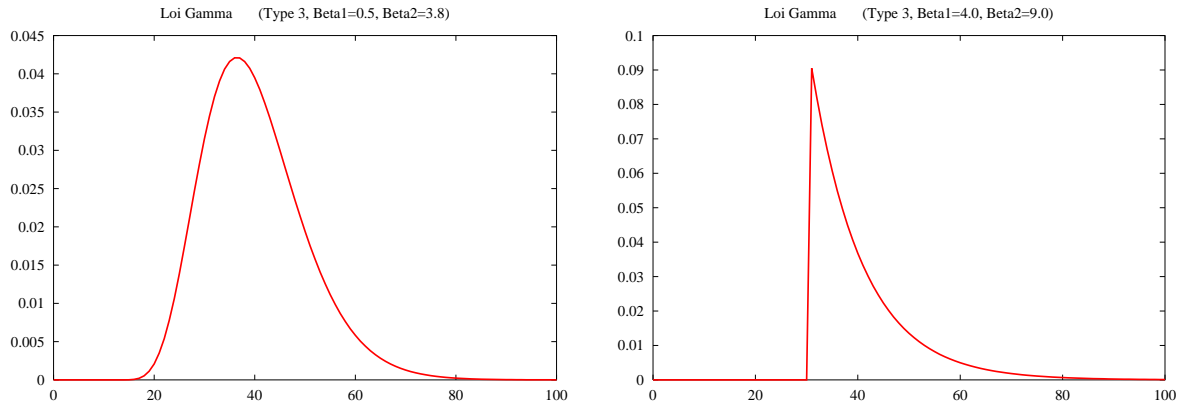


FIG. 4 – Lois gamma.

1.4.3 Type I (lois Béta de 1^{re} espèce)

La famille de lois de Type I intervient lorsque les racines du polynôme (r_1, r_2) sont réelles et de signes opposés. Les conditions sont précisément :

- $\gamma_1 \neq 0$ (sinon cas particulier du type II, cf. section 1.4.4), et
- $-1 < \Delta < 0$ et $\Delta \neq -\frac{1}{2}$ (si $\Delta = -\frac{1}{2}$, alors cas particulier du type XII), et enfin
- $m_1 \neq 0$ ou $m_2 \neq 0$ (cf. ci-dessous) (sinon cas particuliers des types IX et VIII respectivement).

La densité de probabilité et la fonction de répartition s'écrivent :

$$p_1(x) = \frac{1}{r_2 - r_1} f_{m_1+1, m_2+1}^3 \left(\frac{x - r_1}{r_2 - r_1} \right) \quad (12)$$

$$P_1(x) = F_{m_1+1, m_2+1}^3 \left(\frac{x - r_1}{r_2 - r_1} \right) \quad (13)$$

avec

$$r_1 = -\frac{1}{2} \left(\frac{c_1}{c_2} + \frac{\sqrt{\delta}}{|c_2|} \right), \quad r_2 = \frac{1}{2} \left(-\frac{c_1}{c_2} + \frac{\sqrt{\delta}}{|c_2|} \right) \quad (14)$$

$$m_1 = \frac{c_1 + r_1}{c_2 (r_2 - r_1)}, \quad m_2 = -\frac{c_1 + r_2}{c_2 (r_2 - r_1)} \quad (15)$$

pour $r_1 < x < r_2$ ⁷.

La génération de nombres aléatoires est réalisée grâce à :

$$Va_1 = (r_2 - r_1) \text{ gsl_ran_beta}(R, m_1 + 1, m_2 + 1) + r_1 + \mu_1$$

Les paramètres utilisés dans la figure 5 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Beta 1	+40.00	+100.00	+1224.74	+26000.00	+1.50	+2.60
Type XII	+40.00	+100.00	+1224.74	+36000.00	+1.50	+3.60
Beta 1	+40.00	+100.00	+1224.74	+40000.00	+1.50	+4.00
Beta 1	+40.00	+100.00	+1224.74	+49000.00	+1.50	+4.90

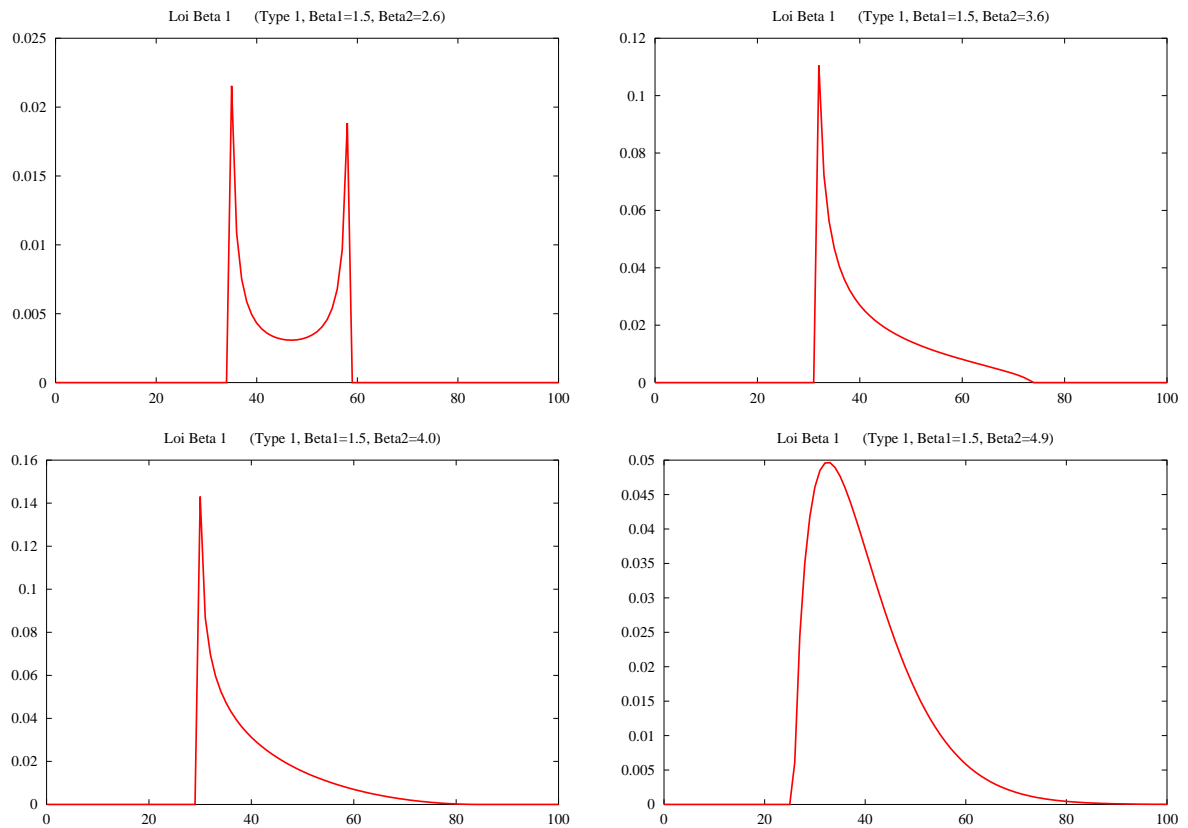


FIG. 5 – Lois beta de 1^{re} espèce.

Cas particuliers : Il existe trois cas particuliers : Type II, Type XII et Types IX et VIII. Ces trois cas s'expriment comme des lois de Type I avec des valeurs de paramètres spécifiques.

- Type XII : Il intervient lorsque $\Delta = -\frac{1}{2}$ (droite d'équation $\beta_2 = \frac{3}{5}(2\beta_1 + 3)$). Dans ce cas, nous avons $m_1 = -m_2$. Ce cas ne pose pas de problème sauf lorsque, en plus des conditions, nous avons $\beta_1 = 0$, c'est à dire lorsque le type XII intersecte le type II.
- Type II : cf. section 1.4.4.
- Types IX et VIII : Ils interviennent lorsque m_1 ou m_2 est nul.

1.4.4 Type II (cas particulier des lois Béta de 1^{re} espèce)

La famille de lois de Type II est un cas particulier de la famille de lois Béta de 1^{re} espèce. Nous l'obtenons lorsque $\gamma_1 = \beta_1 = 0$ (soit $c_1 = 0$) et $\beta_2 \neq \frac{9}{5}$. On retrouve alors $r = -r_1 = r_2$ et $m = m_1 = m_2$.

⁷— r_1 désigne toujours la plus petite des deux racines.

La densité de probabilité et la fonction de répartition s'écrivent :

$$p_2(x) = \frac{1}{2r} f_{m+1,m+1}^3 \left(\frac{x+r}{2r} \right) \quad (16)$$

$$P_2(x) = F_{m+1,m+1}^3 \left(\frac{x+r}{2r} \right) \quad (17)$$

avec

$$r = \frac{\sqrt{\delta}}{2|c_2|} \quad (18)$$

$$R = 2r \quad (19)$$

$$m = -\frac{r}{Rc_2} = -\frac{1}{2c_2} \quad (20)$$

pour $-r < x < r$ ⁸.

La génération de nombres aléatoires est réalisée grâce à :

$$Va_2 = (2r) \text{ gsl_ran_beta}(R, m+1, m+1) - r + \mu_1$$

Les paramètres utilisés dans la figure 6 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Type II	+40.00	+100.00	+0.00	+15000.00	+0.00	+1.50
Type II	+40.00	+100.00	+0.00	+25000.00	+0.00	+2.50

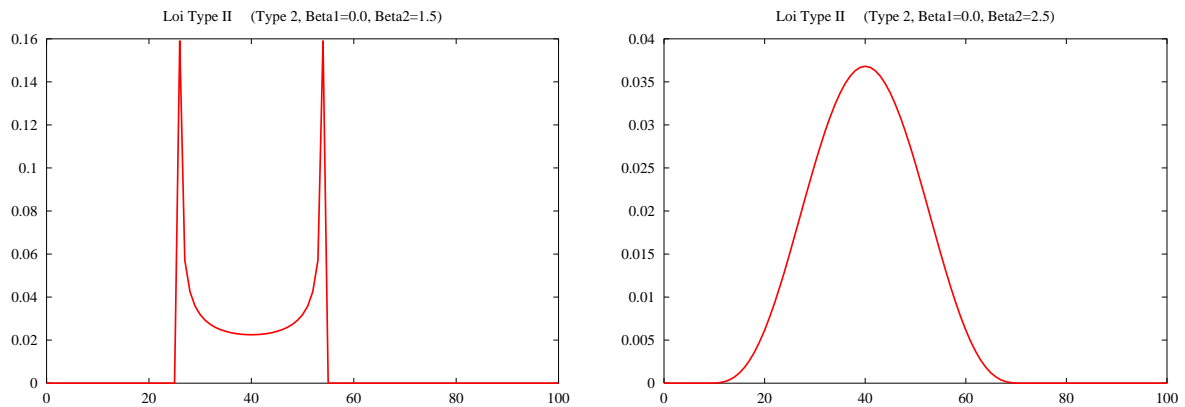


FIG. 6 – Loïs de Type II.

1.4.5 Type VI (lois Béta de 2^e espèce)

Les racines sont de même signe (toutes les deux positives ou toutes les deux négatives). La densité de probabilité s'écrit :

$$p_6(x) = C_6 f_{p_1,p_2}^3 \left(\frac{r_2 - r_1}{x - r_1} \right)$$

pour $\gamma_1 > 0$ et $x > r_2$ ou $\gamma_1 < 0$ et $x < r_2$, avec :

$$r_1 = -\frac{1}{2c_2} \left(c_1 + \text{sign}(\gamma_1) \sqrt{\delta} \right), \quad r_2 = \frac{1}{2c_2} \left(-c_1 + \text{sign}(\gamma_1) \sqrt{\delta} \right) \quad (21)$$

$$R = r_2 - r_1 = -\text{sign}(\gamma_1) \frac{\sqrt{\delta}}{c_2}, \quad p_1 = -m_1 - m_2 + 1, \quad p_2 = m_2 + 1 \quad (22)$$

$$C_6 = \frac{(m_1 + m_2) (m_1 + m_2 + 1)}{(m_1 - 1) m_1 |R|} \quad (23)$$

⁸— r désigne toujours un nombre strictement positif.

avec m_1 et m_2 définis de manière identique à la famille de lois Beta de 1^{re} espèce. Notons également que pour $\gamma_1 > 0$, nous avons $r_1 < r_2 < 0, R > 0$ et pour $\gamma_1 < 0$, nous avons $0 < r_2 < r_1, R < 0$. La fonction de répartition s'écrit :

$$\text{Pour } \gamma_1 > 0, \quad P_6(x) = 1 - F_{p_1-2, p_2}^3 \left(\frac{r_2 - r_1}{x - r_1} \right)$$

$$\text{Pour } \gamma_1 < 0, \quad P_6(x) = F_{p_1-2, p_2}^3 \left(\frac{r_2 - r_1}{x - r_1} \right)$$

La génération de nombres aléatoires est réalisée grâce à :

$$Va_6 = \frac{r_2 - r_1}{gsl_ran_beta(R, p_1 - 2, p_2)} + r_1 + \mu_1$$

Les paramètres utilisés dans la figure 7 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Beta 2	+40.00	+100.00	+707.11	+38000.00	+0.50	+3.80
Beta 2	+40.00	+100.00	+1870.83	+107000.00	+3.50	+10.70
Beta 2	+40.00	+100.00	+1870.83	+83000.00	+3.50	+8.30

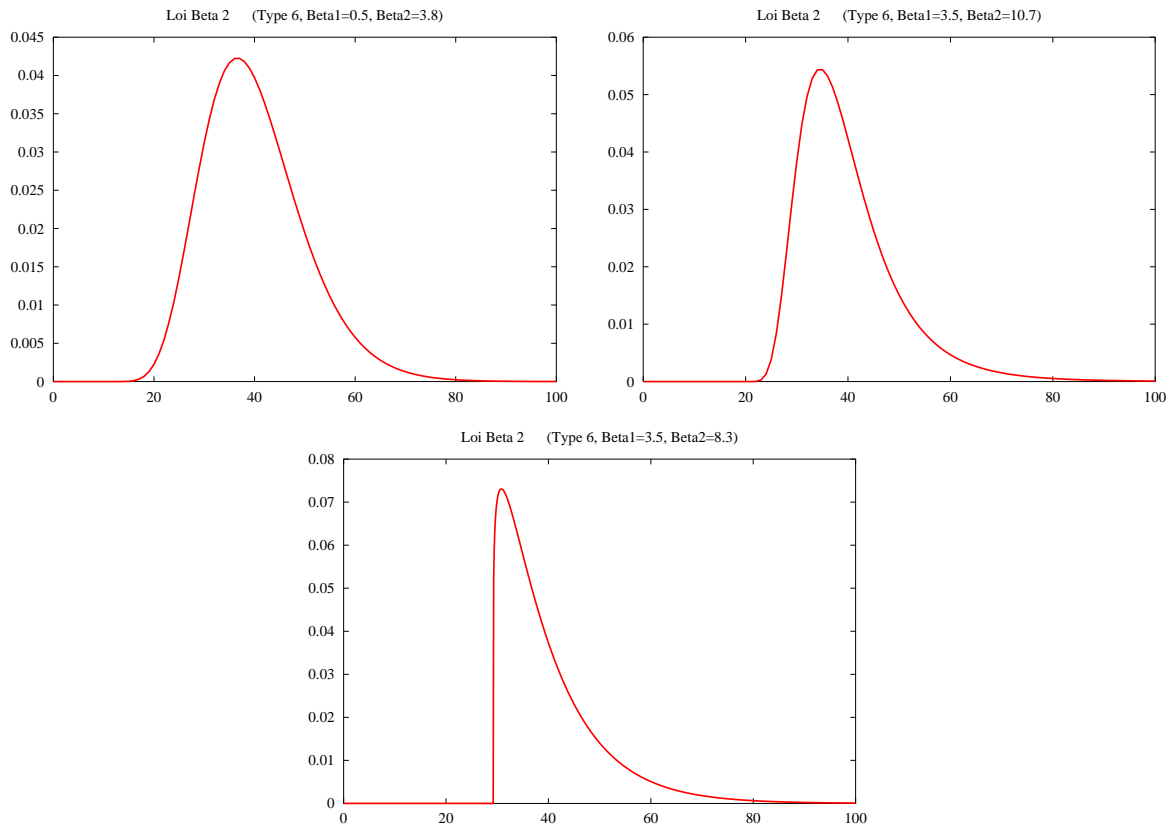


FIG. 7 – Lois Beta de 2^e espèce.

1.4.6 Type IV

La densité de probabilité s'écrit :

$$p_4(x) = C_4 \left((x+r)^2 + S^2 \right)^{-m} e^{-v \tan^{-1} \left(\frac{x+r}{S} \right)} \quad (24)$$

Les coefficients sont donnés par :

$$D = 4\Delta(\Delta + 2) - \beta_1, \quad m = \frac{1 + 2\Delta}{\Delta}, \quad (25)$$

$$v = -2\gamma_1 \frac{1 + \Delta}{\Delta \sqrt{D}}, \quad S = \sigma \frac{\sqrt{D}}{2\Delta}, \quad r = \frac{\gamma_1 \sigma}{2\Delta} \quad (26)$$

$$C_4 = S^{2m-1} \frac{\Gamma(m + \frac{v}{2}i) \Gamma(m - \frac{v}{2}i)}{\Gamma(m) \Gamma(m - \frac{1}{2}) \sqrt{\pi}}. \quad (27)$$

La forme analytique du coefficient C_4 provient des travaux de Y. Nagahara [8]. Les fonctions gamma utilisées dans le numérateur sont complexes conjugué, leur produit est donc un nombre réel de la forme ρ^2 . Le calcul a été implémenté dans GSL sous le nom `gsl_sf_lngamma_complexe_e(...)`.

La fonction de répartition n'a pas de forme analytique évidente et n'a pas été implémentée. De même, la génération de nombres aléatoires n'a pas été implémentée.

Les paramètres utilisés dans la figure 8 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Type IV	+40.00	+100.00	+707.11	+47500.00	+0.50	+4.75
Type IV	+40.00	+100.00	+1732.05	+97000.00	+3.00	+9.70

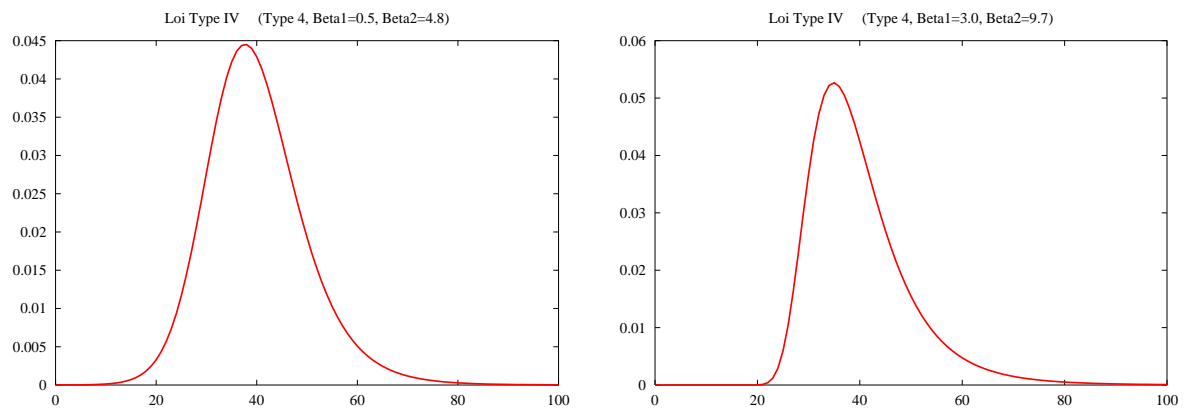


FIG. 8 – Lois Type IV.

1.4.7 Type V (lois Gamma inverse)

Les lois de type V (lois gamma inverse, cf. [9], page 79) font l'interface entre les lois de type IV et de type VI. Dans ce cas, la racine du polynôme est double, ce qui donne : $\delta = 0$, $c_1^2 = 4c_0c_2$. Après quelques calculs, la densité de probabilité s'écrit, en notant $r = -\frac{\sigma\gamma_1}{2\Delta}$:

$$p_5(x) = |C_5| f_{\alpha,1}^2\left(\frac{\theta}{x-r}\right)$$

pour $\gamma_1 > 0$ et $x > r$ ou $\gamma_1 < 0$ et $x < r$, avec :

$$m = \frac{1}{2c_2}, \quad \alpha = 2m + 1, \quad \theta = 2r(1 - m) \quad (28)$$

$$C_5 = \frac{m}{r} \frac{2m - 1}{m - 1} \quad (29)$$

La fonction de répartition s'écrit :

$$\begin{aligned} \text{Pour } \gamma_1 > 0, \quad P_5(x) &= 1 - \Gamma_{\alpha-2} \left(\frac{\theta}{x-r} \right) \\ \text{Pour } \gamma_1 < 0, \quad P_5(x) &= \Gamma_{\alpha-2} \left(\frac{\theta}{x-r} \right) \end{aligned}$$

La génération de nombres aléatoires est réalisée grâce à⁹ :

$$Va_5 = \frac{1}{gsl_ran_gamma(R, \alpha - 2, \beta)} + r + \mu_1$$

Les paramètres utilisés dans la figure 9 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Type V	+40.00	+100.00	+707.11	+39611.32	+0.50	+3.96
Type V	+40.00	+100.00	+1732.05	+95214.33	+3.00	+9.52

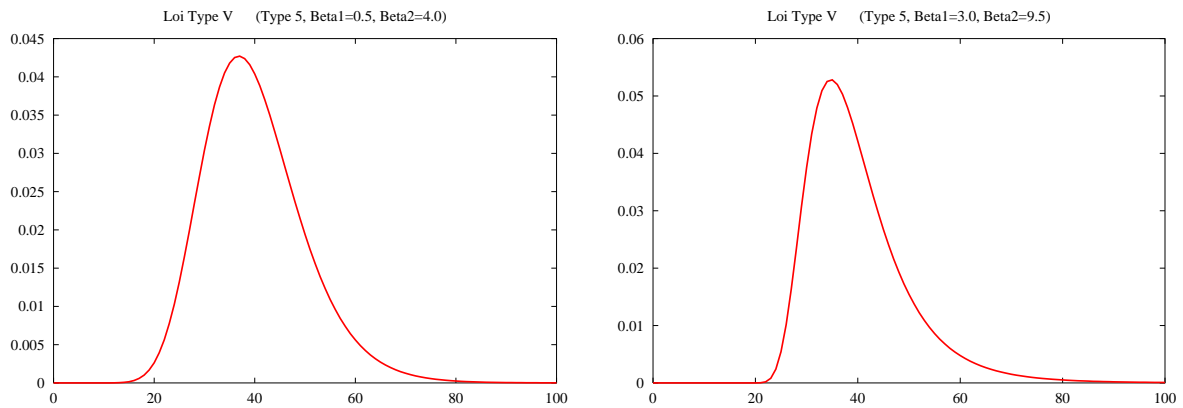


FIG. 9 – Lois de type V.

1.4.8 Type VII (lois de Student't)

La loi de Type VII peut être considéré comme un cas particulier de la loi de Type IV. Elle est traitée de manière séparée car les simplifications de calculs permettent de dégager une forme d'une loi de Student't, de paramètre ν .

Nous avons $\beta_1 = 0$ et $3 < \beta_2 < \frac{9}{2}$ et nous pouvons alors écrire :

$$\begin{cases} c_0 = \mu_2 (1 - 3c_2) = \mu_2 \frac{2\beta_2}{5\beta_2 - 9} \\ c_1 = 0 = a \\ c_2 = \frac{\beta_2 - 3}{5\beta_2 - 9} \end{cases}$$

Remarquons que $c_0 > 0$ et $0 < c_2 < \frac{1}{9}$. L'équation différentielle s'écrit :

$$\frac{d \ln p_7(x)}{dx} = - \frac{x}{c_0 + c_2 x^2}$$

⁹— Pour γ_1 positif ou négatif.

La solution s'écrit finalement :

$$p_7(x) = C_7 \left(1 + \frac{c_2}{c_0} x^2 \right)^{-\frac{1}{2} \frac{1}{c_2}}$$

avec C_7 tel que $\int p_7(x) dx = 1$. En effectuant le changement de variable $y^2 = \frac{1-c_2}{c_0} x^2$, nous obtenons :

$$p_7(x) = C_7 \left(1 + \frac{y^2}{\frac{1-c_2}{c_2}} \right)^{-\frac{1}{2} \frac{1}{c_2}}$$

Cette dernière équation est de la forme d'une loi de Student't dont l'expression est donnée par :

$$S_\nu(t) = \frac{1}{\sqrt{\nu} \beta \left(\frac{1}{2}, \frac{1}{2}\nu\right)} \left(1 + \frac{t^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)}$$

Ainsi, nous obtenons finalement :

$$p_7(x) = C_7 S_\nu(C_7 x).$$

en posant $\nu = \frac{1-c_2}{c_2} > 0$ et

$$C_7 = \sqrt{\frac{1-c_2}{c_0}}$$

La fonction de répartition s'écrit :

$$\begin{cases} \forall x < 0, & P_7(x) = \frac{1}{2} I_{\frac{1}{2}\nu, \nu} \left(\frac{\nu}{\nu + (C_7 x)^2} \right) \\ \forall x \geq 0, & P_7(x) = 1 - \frac{1}{2} I_{\frac{1}{2}\nu, \nu} \left(\frac{\nu}{\nu + (C_7 x)^2} \right) \end{cases}$$

avec la fonction I (fonction bêta incrémentale normalisée) définie dans le tableau 3.

La génération de nombres aléatoires est réalisée grâce à :

$$Va_7 = \frac{1}{C_7} gsl_ran_tdist(R, \nu) + \mu_1$$

Les paramètres utilisés dans la figure 10 sont les suivants :

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Student't	+40.00	+100.00	+0.00	+31000.00	+0.00	+3.10
Student't	+40.00	+100.00	+0.00	+44000.00	+0.00	+4.40

1.5 Illustrations des formes dans le système de Pearson

Les dessins de la figure 11 représentent un panel de formes des principales familles de lois du système de Pearson pour, respectivement, $\gamma_1 > 0$ et $\gamma_1 < 0$ (les lois telles que $\gamma_1 = 0$ restent inchangées dans les deux dessins).

Chaque point de la figure 12 représente la position (β_1, β_2) des lois précédentes dans le diagramme de Pearson. Ce diagramme est identique pour $\gamma_1 > 0$ et $\gamma_1 < 0$ car $\beta_1 = \gamma_1^2$.

Les moments de chacune des lois sont donnés dans le tableau 1¹⁰.

2 Annexes

2.1 Expressions des lois usuelles

cf. table 2

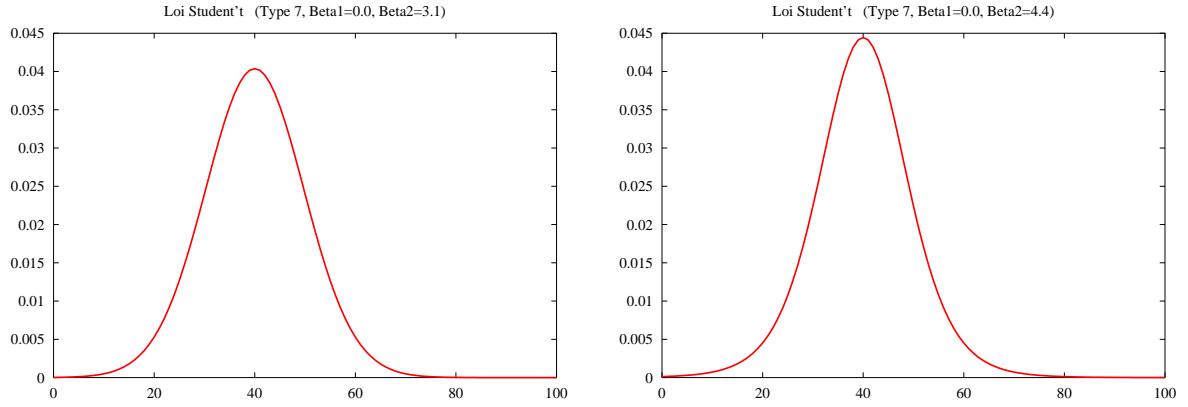


FIG. 10 – Lois de Student't.

TAB. 1 – Illustration de la forme de chaque famille de loi du système de Pearson, pour $\gamma_1 > 0$.

Moments	μ_1	μ_2	μ_3	μ_4	β_1	β_2
Gaussienne	100.00	80.00	0.00	19200.00	0.00	3.00
Gamma	125.00	80.00	1239.35	48000.00	3.00	7.50
Type II	150.00	40.00	0.00	3520.00	0.00	2.20
Beta1	175.00	200.00	1549.19	96000.00	0.30	2.40
Beta 2	200.00	40.00	299.33	8960.00	1.40	5.60
Type IV	225.00	500.00	19364.92	2500000.00	3.00	10.00
Student't	250.00	500.00	0.00	975000.00	0.00	3.90
Type V	275.00	200.00	4898.98	380857.32	3.00	9.52

2.2 Quelques fonctions spéciales

cf. table 3

Références

- [1] N.L. Johnson and S. Kotz. *Continuous univariate distribution*. Tome I. Wiley-interscience, 1970.
- [2] N.L. Johnson and S. Kotz. *Continuous univariate distribution*. Tome II. Wiley-interscience, 1970.
- [3] D.W. Miller. *Fitting frequency distributions - Philosophy and practice*. Part II - Continuous distributions. Book Resource (service@bookresource.com), 2 edition, 1998.
- [4] A. Stuart and J.K. Ord. *Kendall's advanced theory of statistics*. Tome I, Distribution theory. Charles Griffin & Compagny Limited, London, fifth edition, 1986.
- [5] A. Marzouki. *Segmentation statistique d'images radar*. Thèse de 3^e cycle, Université de Lille 1, France, 1996.
- [6] Y. Delignon, A. Marzouki, and W. Pieczynski. Estimation of generalized mixtures and its application in image segmentation. *IEEE trans. on Image Processing*, 6(10) :1364–1375, octobre 1997.

¹⁰— Pour $\gamma_1 < 0$, il suffit de changer le signe de μ_3 .

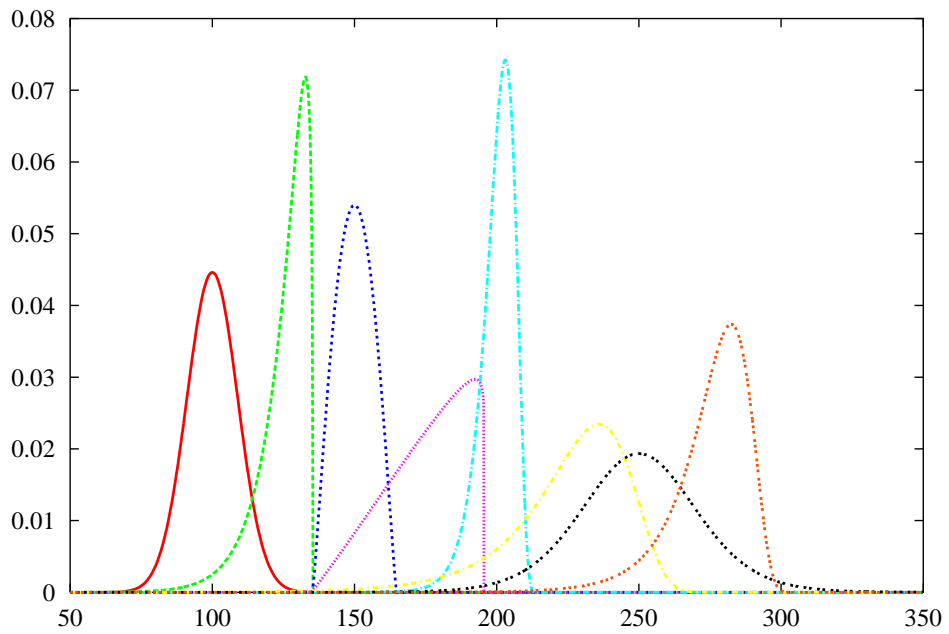
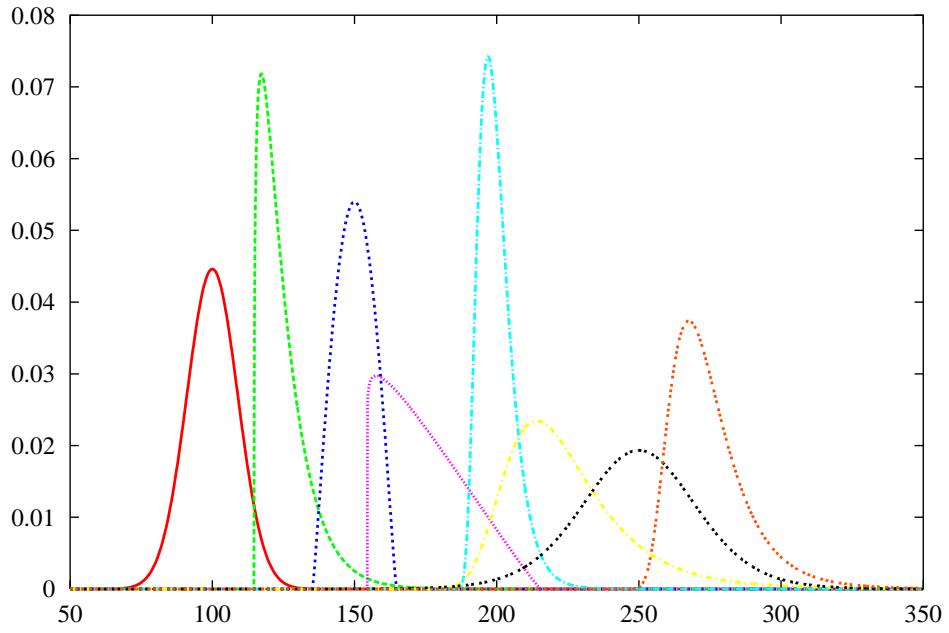


FIG. 11 – En haut : $\gamma_1 > 0$; En bas : $\gamma_1 < 0$.

TAB. 2 – Expression des densités de probabilité et des fonctions de répartition (centrées) des lois usuelles.

Lois	Densité de proba.	Fonction de répartition
Gaussienne $x \in \mathbb{R}$ $\sigma > 0, \mu \in \mathbb{R}$	$f_{\sigma}^1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{x^2}{2\sigma^2}\right)}$	$F_{\sigma}^1(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{\left(-\frac{x^2}{2\sigma^2}\right)} dx = \frac{1}{2} \text{Erfc} \left(\frac{y}{\sigma\sqrt{2}} \right)$
Gamma $x \geq 0$ $\alpha, \theta > 0$	$f_{\alpha, \theta}^2(x) = \frac{1}{\theta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{\left(-\frac{x}{\theta}\right)}$	$F_{\alpha, \theta}^2(y) = \frac{1}{\theta^{\alpha} \Gamma(\alpha)} \int_0^y x^{\alpha-1} e^{\left(-\frac{x}{\theta}\right)} dx = \Gamma_{\alpha} \left(\frac{y}{\theta} \right)$
Beta 1^{er} esp. $x \in [0; 1]$ $a_1, b_1 > 0$	$f_{a_1, b_1}^3(x) = \frac{1}{\beta(a_1, b_1)} x^{a_1-1} (1-x)^{b_1-1}$	$F_{a_1, b_1}^3(y) = \frac{1}{\beta(a_1, b_1)} \int_0^y x^{a_1-1} (1-x)^{b_1-1} dx = \beta_{a_1, b_1}(y)$
Beta 2^e esp. $x \geq 0$ $a_2, b_2 > 0$	$f_{a_2, b_2}^4(x) = \frac{1}{\beta(a_2, b_2)} x^{a_2-1} (1+x)^{-a_2-b_2}$ $= \frac{\beta(a_2, b_2+2)}{\beta(a_2, b_2)} f_{a_2, b_2+2}^3 \left(\frac{x}{x+1} \right)$	$F_{a_2, b_2}^4(y) = \frac{\beta(a_2, b_2+2)}{\beta(a_2, b_2)} \int_0^y f_{a_2, b_2+2}^3 \left(\frac{x}{x+1} \right) dx$ $= \beta_{a_2, b_2} \left(\frac{y}{y+1} \right)$

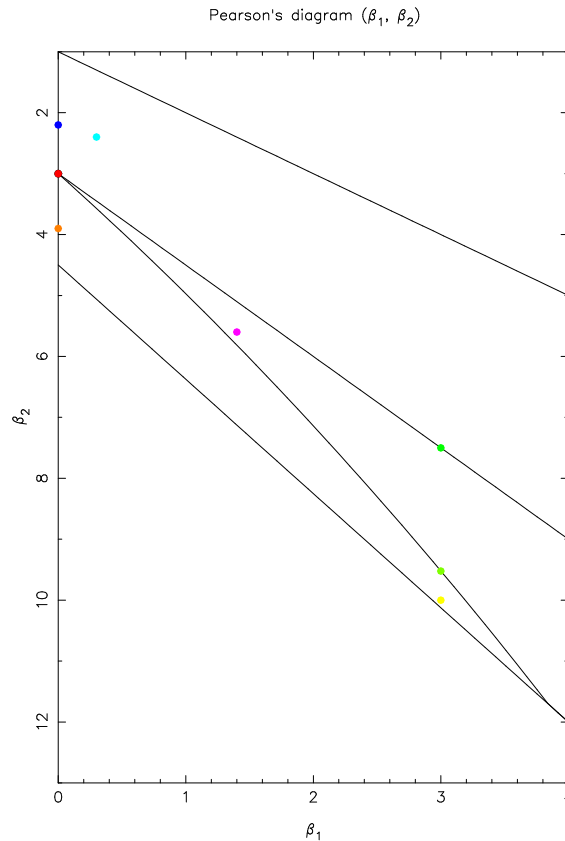


FIG. 12 – Caractérisation des lois de la figure 11 dans le diagramme de Pearson.

- [7] S. Colombo. *Les transformations de Mellin et de Hankel*. Centre National de la Recherche Scientifique, 1959.
- [8] Y. Nagahara. The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. *Statistical & Probability Letters*, 43 :251–264, 1999.
- [9] M. Evans, N. Hastings, and B. Peacock. *Statistical distributions*. Wiley-interscience, 2 edition, 1983.
- [10] G. Kaskey, P.R. Krishnaiah, Kolman B., and L. Steinberg. *Handbook of Statistics*, volume 1, chapter Transformation to normality, pages 321–341. North-Holland Publishing Company, 1980.
- [11] Y. Delignon. *Études statistiques d'images radar de la surface de la mer*. Thèse de 3^e cycle, Université de Rennes 1, France, 1993.
- [12] N. Unnikrishnan Nair and P.G. Sankaran. Characterization of the Pearson family of distributions. *IEEE trans. on Reliability*, 40(1) :75–77, april 1991.
- [13] R.S. Parrish. On an integrated approach to member selection and parameter estimation for Pearson distributions. *Computational Statistics & Data Analysis*, 1 :239–255, 1983.
- [14] W. Albers, W.C.M. Kallenberg, and G.D. Otten. A comparison between Rosenblatt's estimator and parametric density estimators for determining test limits. *Computational Statistics & Data Analysis*, 27 :47–60, 1998.
- [15] S. Medasani and R. Krishnapuram. A comparison of Gaussian and Pearson mixture modeling for pattern recognition and computer vision applications. *Pattern Recognition Letters*, 20 :305–313, 1999.

TAB. 3 – Quelques fonctions spéciales intervenant dans le calcul des densités de probabilité.

Erf	$Erf(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^x e^{-t^2} dt$
Erfc	$Erfc(x) = 1 - Erf(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$
Gamma	$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha-1} dt, \alpha > 0$
Gamma inc.	$\Gamma_{\alpha}(x) = \frac{1}{\Gamma(\alpha)} \int_0^x e^{-t} t^{\alpha-1} dt, x \geq 0, \alpha > 0$
DiGamma	$\Psi(x) = \frac{d \ln(\Gamma(x))}{dx} = \frac{1}{\Gamma(x)} \frac{d\Gamma(x)}{dx}$
Beta	$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \alpha, \beta > 0$
Beta inc.	$\beta_{a,b}(x) = \frac{1}{\beta(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, x \in [0; 1], \alpha, \beta > 0$
Beta inc. norm.	$I_{a,b}(x) = \frac{\beta_{a,b}(x)}{\beta_{a,b}}, x \in [0; 1], \alpha, \beta > 0$

- [16] S. Chitroub, A. Houacine, and B. Sansal. Statistical characterization and modelling of SAR images. *Signal Processing*, 82 :69–92, 2002.
- [17] J. Karvanen and V. Koivunen. Blind separation methods based on Pearson system and its extensions. *Signal Processing*, ?? :??–??, 2002.
- [18] S. Derrode, G. Mercier, J-M. Le Caillec, and R. Garello. Estimation of sea-ice SAR clutter statistics from Pearson’s system of distributions. In *IEEE Int. Conf. on Geoscience and Remote Sensing (IGARSS’01)*, Sydney (Australia), 9-13 Juillet 2001.
- [19] J.M. Nicolas, M. Sigelle, C. Thuillier, and F. Tupin. Images de radar à ouverture synthétique : transformée de mellin et multirésolution. In *GRETSI’97*, pages 797–800, Grenoble (France), Septembre 1997.