Lois de variables aléatoires continues

6.1 Généralités

Une variable aléatoire continue prend ses valeurs sur un ensemble infini non dénombrable de points, elle décrit par exemple la durée de vie d'une batterie de voiture, l'heure d'arrivée des voitures à un péage donné d'autoroute...

Il existe une fonction f non négative, définie pour toute valeur x appartenant à $\mathbb{R}$ et vérifiant, pour toute partie A de $\mathbb{R}$, la propriété :

$$\Pr(X \in A) = \int_A f(x)\,\mathrm{d}x$$

et telle que :

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$$

La fonction f est la ***densité de probabilité*** de la variable aléatoire X.

La *fonction de répartition* de la variable aléatoire X est définie par :

$$F(a) = \Pr(X < a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x$$

Pour toutes les valeurs a et b appartenant à $\mathbb{R}$, on a donc la relation :

$$\Pr(a \leqslant X < b) = F(b) - F(a)$$

On en déduit :

$$\Pr(X = x) = 0 \qquad \Pr(x \leqslant X < x + \mathrm{d}x) = f(x)\,\mathrm{d}x$$

Espérance mathématique :

$$E(X) = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x$$

Variance :

$$\mathrm{Var}(X) = E\left(X^2\right) - \left[E(X)^2\right] = \int_{\mathbb{R}} x^2 f(x)\,\mathrm{d}x - \left[E(X)^2\right]$$

L'espérance et la variance existent si les intégrales sont définies.

La plupart des problèmes rencontrés en statistique peuvent se résoudre à l'aide de quelques lois fondamentales continues, environ une dizaine.

Les principales sont la loi uniforme, la loi exponentielle, les lois gamma, les lois bêta, la loi normale et la loi log-normale auxquelles il faut ajouter les lois du chi-deux, de Fisher-Snedecor et de Student utilisées dans la théorie de l'estimation (lois étudiées, chapitre 10, paragraphes 10.4, 10.5 et 10.6).

6.2 Loi uniforme

6.2.1 Définition

Une variable aléatoire réelle X, suit une *loi uniforme sur l'intervalle* $[a,\ b]$, si sa loi de probabilité admet *une densité* f *égale à* :

$$f(x) = \frac{1}{b-a}\, 1_{[a,\ b]}$$

$1_{[a,\ b]}$ est la fonction caractéristique du segment $[a,\ b]$.

Fonction de répartition :

$$\begin{aligned} F(x) &= 0 \quad \text{si} \quad x \leqslant a \\ F(x) &= (x-a)/(b-a) \quad \text{si} \quad a < x < b \\ F(x) &= 1 \quad \text{si} \quad x \geqslant b \end{aligned}$$

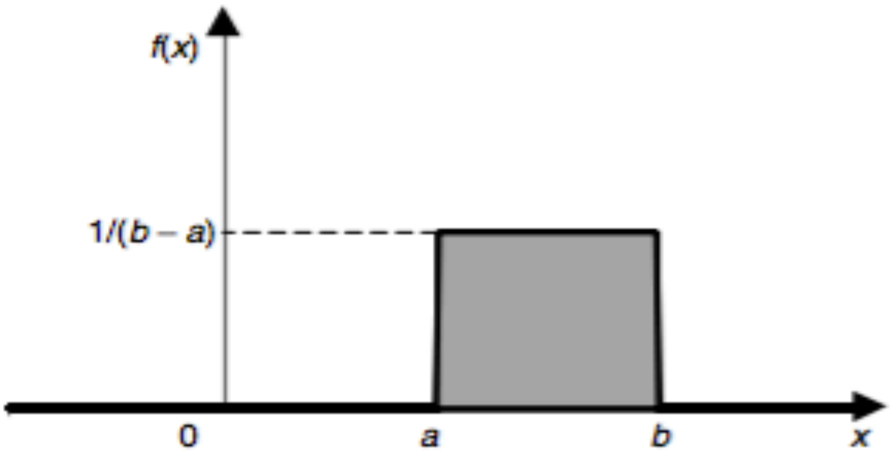

Figure 6.1 – Loi uniforme sur [*a*, *b*]. Densité.

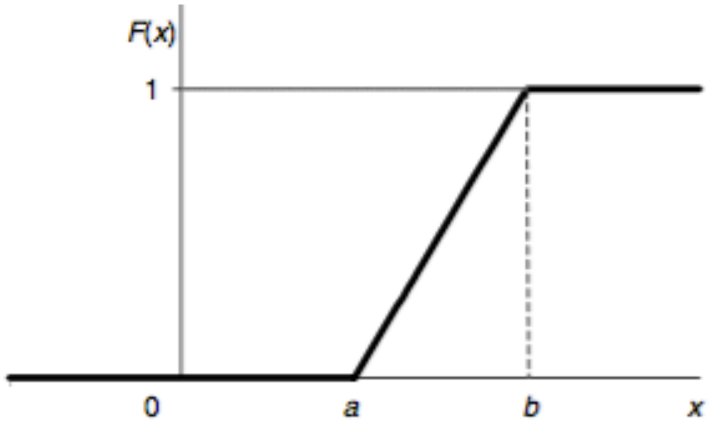

Figure 6.2 – Loi uniforme sur $[a, b]$. Fonction de répartition.

6.2.2 Moments

$$E(X) = \frac{(b+a)}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

6.2.3 Propriétés et domaine d'utilisation

– La somme de deux variables aléatoires, indépendantes ou non, suivant une loi uniforme sur $[a, b]$, ne suit pas une loi uniforme sur $[a, b]$.

– L'image, par sa fonction de répartition, de toute variable aléatoire réelle continue, est une variable aléatoire réelle suivant la loi uniforme sur $[0, 1]$. Cette propriété est utilisée, pour simuler ou engendrer, des échantillons de la loi de la variable aléatoire X (chapitre 11, paragraphe 11.5.4).

Démonstration de cette propriété

Soit X une variable aléatoire dont la fonction de répartition F est continue et strictement croissante.

On considère la variable aléatoire $Y = F(X)$, elle varie de 0 à 1. On désigne par G et g sa fonction de répartition et sa densité :

$$G(y) = \Pr(Y < y) = \Pr(F(X) < y) = \Pr(X < F^{-1}(y)) = F[F^{-1}(y)] = y$$

En résumé :

Si $y \in [-\infty, 0]$

$G(y) = 0$ et $g(y) = 0$

Si $y \in [0, 1]$

$G(y) = y$ et $g(y) = 1$

loi uniforme sur $[0, 1]$.

- La loi uniforme sur $[a, b]$ traduit l'hypothèse d'équirépartition, ou répartition indifférente, sur $[a, b]$. Les tables, concernant ce type de répartition, sont les tables de nombres au hasard (chapitre 11, paragraphe 11.5.3).
- La loi uniforme est utilisée en statistique bayésienne, pour déterminer les lois de probabilité a priori, dans le cas de l'ignorance totale, dans l'intervalle $[0, 1]$ (elle est dite non informative) ou dans l'intervalle $[a, b]$, en utilisant les résultats de l'expert (elle est dite informative).

6.3 Loi exponentielle

6.3.1 Définition

Une variable aléatoire réelle positive X suit une *loi exponentielle, de paramètre λ positif, si sa densité de probabilité* est donnée par :

$$
\begin{aligned}
f(x) &= \lambda e^{-\lambda x} \quad && \text{si } x \geqslant 0 \\
f(x) &= 0 && \text{sinon}
\end{aligned}
$$

X est appelée *variable exponentielle.*

Fonction de répartition :

$$F(a) = \Pr(X < a) = \int_0^a \lambda\, e^{-\lambda x}\, dx = 1 - e^{-\lambda a}$$

6.3.2 Moments

Espérance et variance :

$$E(X) = \frac{1}{\lambda} \qquad \operatorname{Var}(X) = \frac{1}{\lambda^2}$$

L'espérance d'une variable exponentielle est égale à son écart-type.

Coefficients d'asymétrie et d'aplatissement :

$$\gamma_1 = 2 \qquad \gamma_2 = 9$$

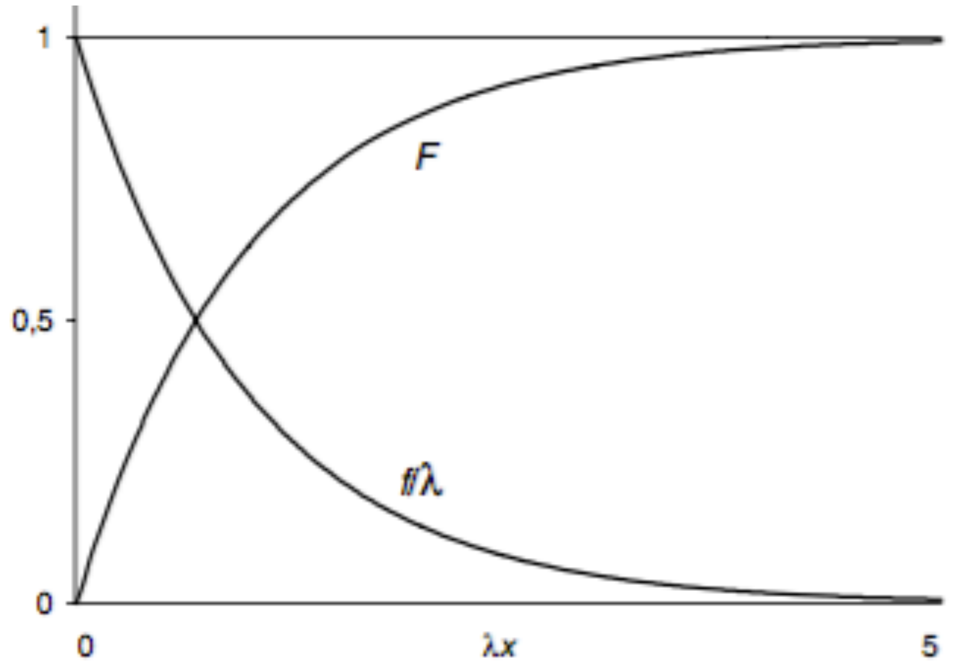

Figure 6.3 – Loi exponentielle. Densité et fonction de répartition de la loi exponentielle (l'axe des abscisses est gradué proportionnellement aux valeurs λx).

6.3.3 Domaine d'utilisation

– La distribution exponentielle est associée aux processus de Poisson. Un tel processus génère des événements dont les temps d'occurrence sont indépendants et distribués suivant une loi exponentielle (chapitre 9).

– La loi exponentielle est utilisée en *fiabilité* (chapitre 18), le paramètre λ représente le taux de défaillance alors que son inverse $\theta = 1/\lambda$ est le temps moyen de bon fonctionnement MTBF (*Mean Time Between Failure*). Avec le paramètre θ, la densité de probabilité s'écrit :

$$f(x) = \frac{1}{\theta}\ e^{-\frac{x}{\theta}}$$

et les moments sont égaux à :

$$E(X) = \theta$$

$$\mathrm{Var}(X) = \theta^2$$

– La loi exponentielle s'applique bien aux matériels électroniques, c'est-à-dire aux matériels fonctionnant pratiquement sans usure, aux matériels subissant des défaillances brutales ou à des systèmes complexes dont les composants ont des lois de fiabilité différentes. Elle permet de décrire la période de fonctionnement durant laquelle le taux de défaillance est constant ou presque constant.

6.3.4 Propriétés

– La somme de deux variables aléatoires indépendantes, suivant des lois exponentielles de paramètres respectifs λ_1 et λ_2, est une variable aléatoire suivant une loi exponentielle de paramètre $\lambda_1 + \lambda_2$.
– La loi exponentielle est qualifiée de loi « sans mémoire », elle permet la modélisation du comportement des matériels fonctionnant avec un taux de défaillance constant (ou pouvant être considéré comme constant).
– On considère un matériel ayant fonctionné sans défaillance pendant le temps x_1 et on cherche la probabilité qu'il soit encore en état de marche au temps $x + x_1$. La définition de la probabilité conditionnelle donne :

$$\begin{aligned}\Pr(X \geqslant x + x_1 \,/\, X \geqslant x_1) &= \frac{\Pr(X \geqslant x + x_1 \text{ et } X \geqslant x_1)}{\Pr(X \geqslant x_1)} \\ &= \frac{e^{-\lambda(x+x_1)}}{e^{-\lambda x_1}} = e^{-\lambda x}\end{aligned}$$

Le matériel a « oublié » qu'il avait déjà fonctionné pendant le temps x_1. Pour ce type de matériel, il est inutile de procéder à un remplacement préventif.

> **Exemple 6.1**
> On suppose que le temps, en heures, nécessaire pour réparer une machine est une variable aléatoire suivant une loi exponentielle de paramètre $\lambda = 0{,}5$.
> La densité de probabilité est $f(t) = 0{,}5\,e^{-0{,}5\,t}$ et la fonction de répartition $F(t) = 1 - e^{-0{,}5\,t}$.
> La probabilité pour que le temps de réparation dépasse 2 heures est :
>
> $$\Pr(T > 2) = 1 - \Pr(T < 2) = 1 - F(2) = e^{-1} = 0{,}368$$
>
> Sachant que la réparation a déjà dépassé 9 heures, quelle est la probabilité qu'elle prenne au moins 10 heures ?
> La loi exponentielle étant une loi sans « mémoire », on obtient :
>
> $$\Pr(T > 10/T > 9) = \Pr(T > 10 - 9 = 1) = e^{-0{,}5} = 0{,}606$$

6.4 Loi gamma

6.4.1 Définition

La loi exponentielle est un cas particulier de la famille des lois gamma.

Une variable aléatoire réelle positive X suit une *loi gamma* $\gamma(t\,;\lambda)$ *ou* $\Gamma(t\,;\lambda)$, *de paramètres positifs t et* λ, *si sa densité de probabilité* est donnée par :

$$f(x) = \frac{\lambda\,e^{-\lambda x}\,(\lambda x)^{t-1}}{\Gamma(t)} \quad \text{si } x \geqslant 0$$

$$f(x) = 0 \quad \text{sinon}$$

Γ est la *fonction eulérienne* définie par l'intégrale pour $t > 0$ (voir annexe 2) :

$$\Gamma(t) = \int_0^{\infty} e^{-y}\,y^{t-1}\,dy$$

Le paramètre t est un paramètre de forme tandis que $1/\lambda$ est un paramètre d'échelle. Pour les représentations graphiques, on peut prendre $\lambda = 1$.

Selon les valeurs du paramètre t, la densité de la loi gamma a différentes formes (figures 6.4 et 6.5).

En particulier, si $t = 1$, on retrouve la loi exponentielle.

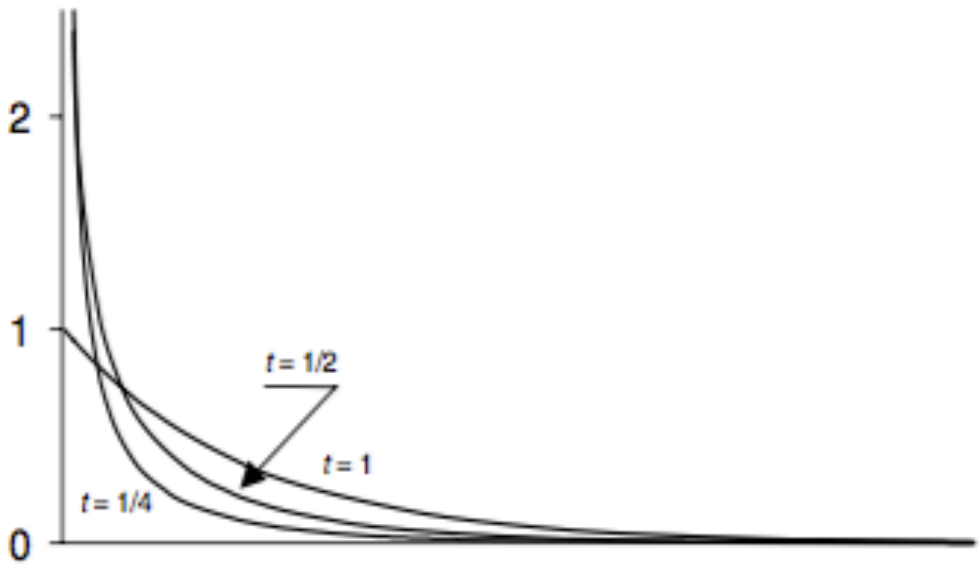

Figure 6.4 – Loi gamma ($t \leqslant 1$).

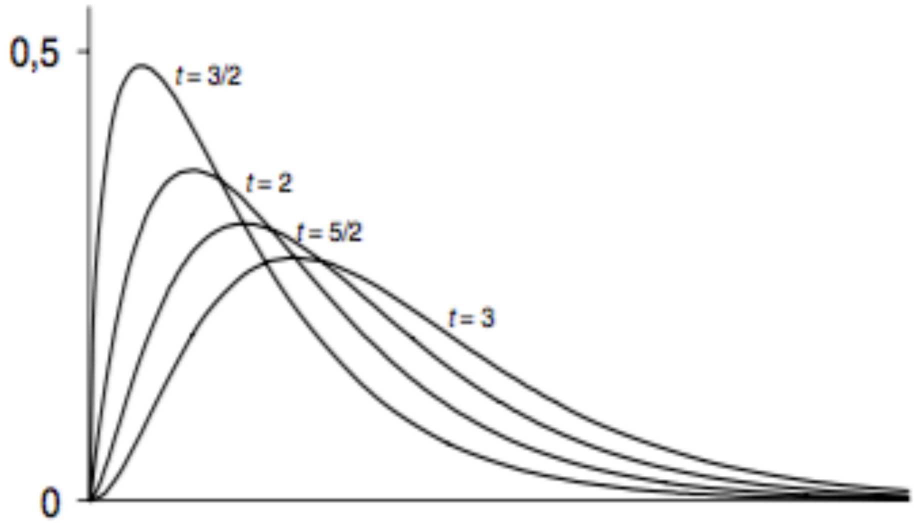

Figure 6.5 – Loi gamma ($t > 1$).

Si le paramètre λ est différent de 1, la variable aléatoire $Y = \lambda X$ suit une loi $\gamma(t\,;1)$ ou $\gamma(t)$ de densité :

$$f(y) = \frac{e^{-y}\, y^{t-1}}{\Gamma(t)} \quad \text{si } y \geqslant 0$$

$$f(y) = 0 \qquad \text{sinon}$$

6.4.2 Moments

Par intégrations par parties et en utilisant les propriétés de la fonction Γ, on obtient :

$$E(X) = \frac{t}{\lambda} \qquad \mathrm{Var}(X) = \frac{t}{\lambda^2}$$

6.4.3 Propriétés et domaine d'utilisation

La loi gamma dépendant de deux paramètres peut être utilisée pour représenter un grand nombre de distributions. Ainsi :

- dans la théorie des files d'attente, la loi gamma représente la loi de probabilité d'occurrence de t événements (t étant un entier), dans un processus poissonnien. Si le temps T, entre les défaillances successives d'un système, suit une loi exponentielle, le temps cumulé d'apparitions de λ défaillances suit une loi gamma $\gamma(t\,;\lambda)$,
- en fiabilité, la loi gamma peut être utilisée pour modéliser les temps de défaillance d'un matériel,

– selon les valeurs des paramètres, la loi gamma s'identifie à d'autres lois :
  - la loi exponentielle si $t = 1$,
  - la loi d'Erlang si t est égal à un entier n supérieur à 1, sa densité est :

$$f(x) = \frac{\lambda\, e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!} \quad \text{si } x \geqslant 0$$

$$f(x) = 0 \quad \text{sinon}$$

    En effet, $\Gamma(n) = (n-1)!$ et sa fonction de répartition, qui correspond à l'apparition de n événements en un temps inférieur à x, est donnée par l'expression :

$$F(x) = 1 - e^{-\lambda x} \sum_{i=1}^{n} \frac{(\lambda x)^{i-1}}{(i-1)!} \quad \text{si } x \geqslant 0$$

$$F(x) = 0 \quad \text{sinon}$$

  - la loi de la variable chi-deux à n degrés de liberté, $\chi^2(n)$, utilisée en statistique, si $\lambda = 1/2$ et $t = n/2$, où n est un entier positif (chapitre 10, paragraphe 10.4),

– la somme de deux variables aléatoires indépendantes, suivant des lois gamma $\gamma(t\,;\lambda)$ et $\gamma(u\,;\lambda)$, suit une loi gamma $\gamma(t + u\,;\lambda)$ (propriété d'additivité des lois gamma).

6.5 Lois bêta de types I et II

6.5.1 Définitions

■ Loi bêta de type I

Une variable aléatoire réelle X, *prenant ses valeurs dans l'intervalle* [0, 1], suit une *loi bêta de type I, notée* $\beta(n\,;p)$, *de paramètres positifs n et p, si sa densité de probabilité* est donnée par :

$$f(x) = \frac{1}{B(n\,;p)}\, x^{n-1} (1-x)^{p-1} \qquad 0 \leqslant x \leqslant 1$$

$$f(x) = 0 \qquad \text{sinon}$$

où $B(n\,;p)$ est la fonction eulérienne définie par (voir annexe 2) :

$$B(n;\ p) = \int_0^1 x^{n-1}\ (1-x)^{p-1}\ \mathrm{d}x = B(p;\ n)$$

$$B(n;\ p) = \frac{\Gamma\ (n)\,\Gamma\ (p)}{\Gamma\ (n+p)}$$

La forme de la densité de X varie selon la valeur des paramètres n et p.

■ Loi bêta de type II

Soit X une variable aléatoire suivant une loi bêta de type I, $\beta(n\,;\,p)$. La variable aléatoire Y, positive ou nulle, définie par $Y = \dfrac{X}{1-X}$, suit une *loi bêta de type II* dont la densité s'obtient facilement par changement de variables :

$$f(y) = \frac{1}{B(n,\ p)}\ \frac{y^{n-1}}{(1+y)^{n+p}} \quad \text{si } y \geqslant 0$$

$$f(y) = 0 \quad \text{sinon}$$

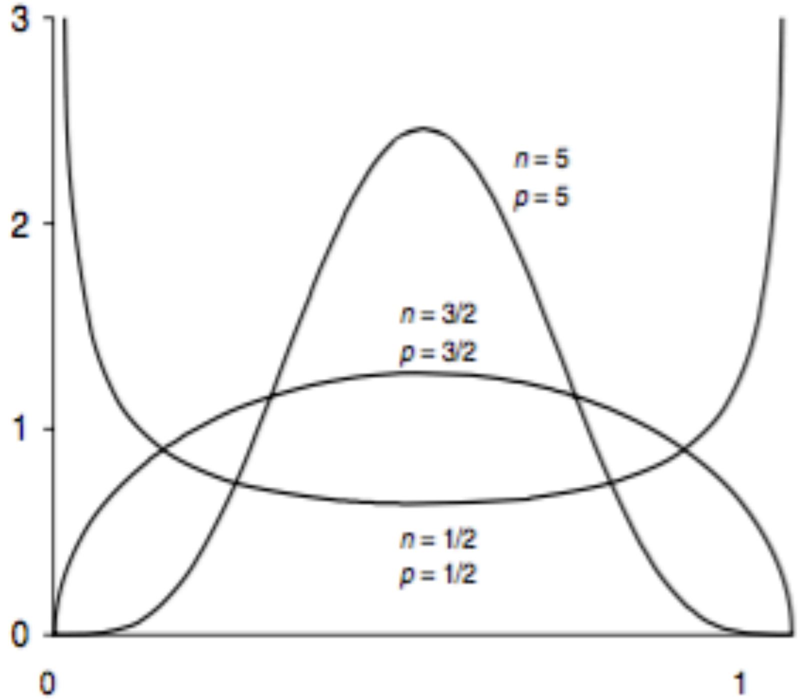

Figure 6.6 – Densité de la loi bêta II, paramètres égaux.

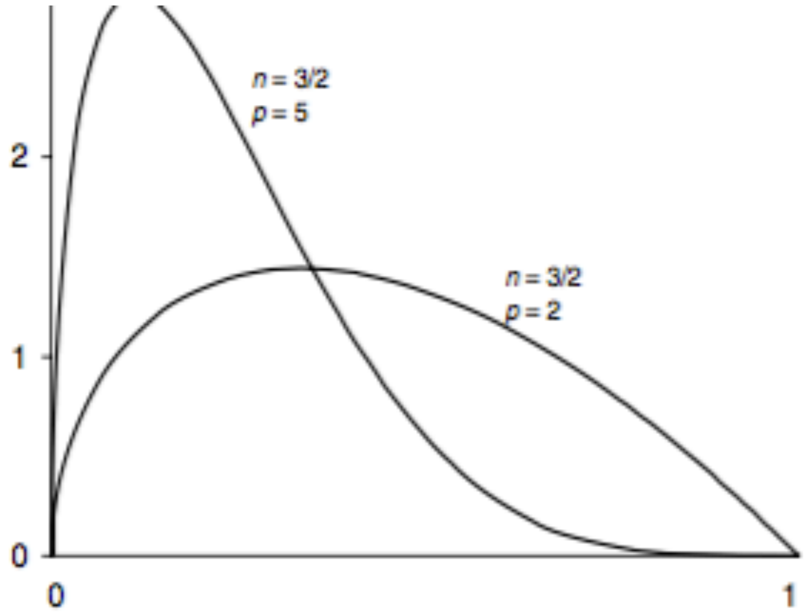

Figure 6.7 – Densité de la loi bêta II, paramètres différents et supérieurs à 1.

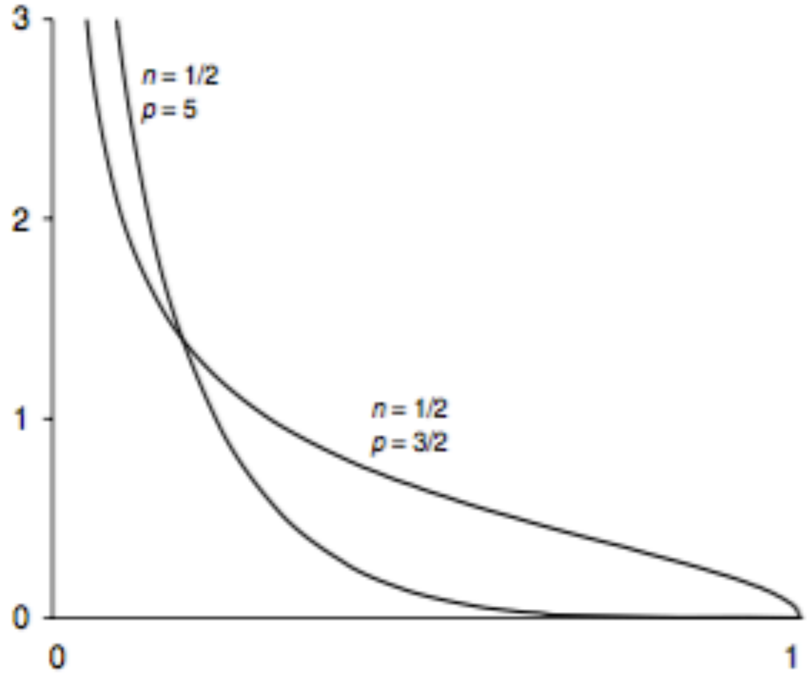

Figure 6.8 – Densité de la loi bêta II, paramètres différents dont un est inférieur à 1.

6.5.2 Moments

■ Loi bêta de type I

$$E(X) = \frac{n}{n+p} \qquad \mathrm{Var}(X) = \frac{np}{(n+p+1)\ (n+p)^2}$$

■ **Loi bêta de type II**

$$E(X) = \frac{n}{p-1} \quad \mathrm{Var}(X) = \frac{n\,(n+p-1)}{(p-1)^2\,(p-2)^2}$$

6.5.3 Propriétés et domaines d'utilisation

- Le *rapport de deux variables aléatoires indépendantes*, suivant les lois gamma $\gamma(t\,;\lambda)$ et $\gamma(u\,;\lambda)$, suit une loi bêta de type II de paramètres t et u.
- Les lois bêta, dépendant de deux paramètres, s'adaptent bien à la description de nombreux phénomènes aléatoires positifs (temps d'attente, durées de vie…) ; elles sont liées aux lois de Fisher-Snedecor utilisées en statistique.
- Les lois bêta de type I sont utilisées en fiabilité, en statistique bayésienne pour représenter la distribution *a priori* de la probabilité d'un événement suivant une loi binomiale, la distribution *a posteriori* suit aussi une loi binomiale.

6.6 Loi de Laplace-Gauss ou loi normale

6.6.1 Définition

Une variable aléatoire réelle X, prenant ses valeurs dans $\mathbb{R}$, suit une *loi de Laplace-Gauss ou loi normale, de paramètres m et* σ, *si sa densité de probabilité* est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}$$

La fonction f définit une densité. En effet :

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

Cette loi est notée, en général $N(m\,;\sigma)$. On dit indifféremment qu'une variable suivant une telle loi est *une variable normale ou gaussienne.*

Fonction de répartition :

$$\Pr(X < a) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{a} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}\,\mathrm{d}x$$

Cette intégrale n'ayant pas d'expression mathématique simple, des tables donnent les valeurs de la fonction de répartition.

Sur la courbe représentant la densité de probabilité d'une variable gaussienne, la valeur de $F(a)$ est représentée par la partie non hachurée. Cette courbe a un axe de symétrie vertical pour $x = m$ et du fait de sa forme, elle est souvent appelée « courbe en cloche ».

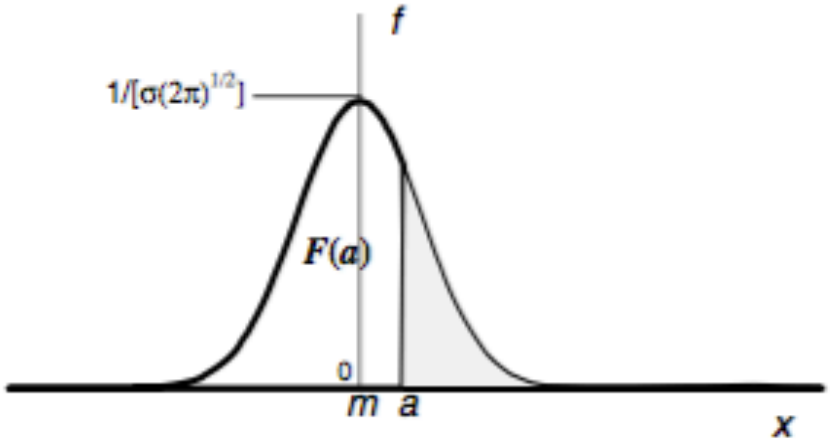

Figure 6.9 – Densité de la loi normale.

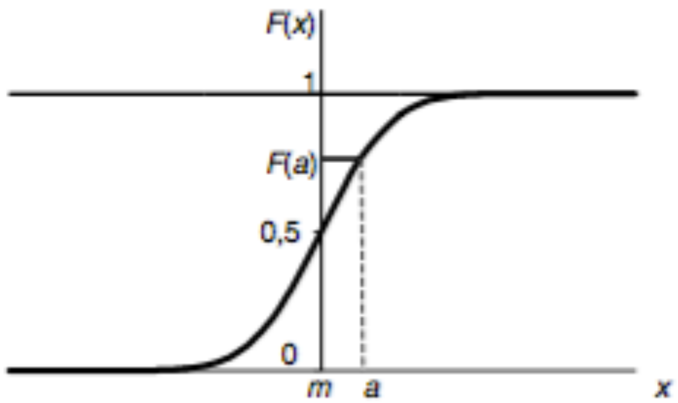

Figure 6.10 – Fonction de répartition de la loi normale.

6.6.2 Moments

Espérance et variance :

$$E(X) = m \qquad \mathrm{Var}(X) = \sigma^2$$

Ces résultats justifient le choix des deux paramètres figurant dans l'expression de la densité.

– Les *moments de tous les ordres existent*. En effet, les intégrales :

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^k e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$$

convergent pour toutes les valeurs de k.

– Les *moments centrés d'ordre impair* sont tous nuls (propriété de symétrie de la densité).

– Les *moments centrés d'ordre pair* ont pour valeurs :

$$\mu_{2k} = 1 \times 3 \times \cdots \times (2k-1)\ \sigma^{2k} = \frac{(2k)!}{2^k\, k!}\sigma^{2k}$$

- $\mu_3 = 0$, le coefficient d'asymétrie γ_1 est nul,
- $\mu_4 = 3\sigma^4$, le coefficient d'aplatissement γ_2 est égal à 3.

6.6.3 Variable aléatoire centrée réduite

La *variable centrée réduite* associée à la variable aléatoire X est la variable :

$$U = \frac{X-m}{\sigma}$$

Ses moments d'ordre impair sont nuls, en particulier $E(U) = 0$ et les moments d'ordre pair sont égaux à :

$$\mu_{2k} = 1 \times 3 \times \cdots \times (2k-1) = \frac{(2k)!}{2^k\, k!}$$

en particulier $\mathrm{Var}(U) = 1$.

Densité de probabilité de la variable U :

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

La variable U suit la loi normale $N(0\,;1)$ dont les paramètres sont $m = 0$ et $\sigma = 1$.

Fonction de répartition :

$$\Pr(X < a) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{u^2}{2}}\,du$$

La table 5.1 donne la fonction de répartition, et la table 5.2 les fractiles de la loi normale réduite car elle ne dépend d'aucun paramètre, les formules de changement de variables :

$$U = \frac{X - m}{\sigma} \text{ et } X = \sigma\, U + m$$

permettant de passer d'une variable à l'autre.

Exemple 6.2 Utilisation de la table de la loi normale

Soit X une variable suivant la loi normale $N(3\,;2)$, donc de moyenne 3 et d'écart-type 2. On veut calculer les probabilités suivantes : $\Pr(X < 4)$, $\Pr(X < -1)$, $\Pr(X > 1)$ ou les nombres a_i tels que $\Pr(X < a_1) = 0{,}75$, $\Pr(X > a_2) = 0{,}85$. On utilise la variable centrée réduite U associée à la variable X :

$$U = \frac{X - 3}{2} \quad \text{et} \quad X = 2\,U + 3$$

$$\Pr(X < 4) = \Pr(2\,U + 3 < 4) = \Pr(U < 0{,}50) = 0{,}6915$$

$$\Pr(X < -1) = \Pr(2\,U + 3 < -1) = \Pr(U < -2) = 0{,}0228$$

$$\Pr(X > 1) = \Pr(2\,U + 3 > 1) = \Pr(U > -1) = 0{,}8413$$

$$\Pr(X < a_1) = \Pr(2\,U + 3 < a_1) = \Pr\left(U < \frac{a_1 - 3}{2}\right) = 0{,}75$$

$\Pr(U < 0{,}6745) = 0{,}75$ D'où : $a_1 = 4{,}35$

$$\Pr(X > a_2) = \Pr(2\,U + 3 > a_2) = \Pr\left(U > \frac{a_2 - 3}{2}\right) = 0{,}85$$

$\Pr(U < -1{,}0364) = 0{,}15$ $\Pr(U > -1{,}0364) = 0{,}85$

D'où : $a_2 = -1{,}0364 \times 2 + 3 = 0{,}9272$

Résultats remarquables :

$$\Pr(m - 1{,}64\sigma < X < m + 1{,}64\sigma) = 0{,}90$$

$$\Pr(m - 1{,}96\sigma < X < m + 1{,}96\sigma) = 0{,}95$$

$$\Pr(m - 3{,}09\sigma < X < m + 3{,}09\sigma) = 0{,}998$$

Loi de la variable U^2 : la densité de probabilité g de la variable aléatoire $T = U^2$ est obtenue en utilisant la formule donnant la densité de probabilité