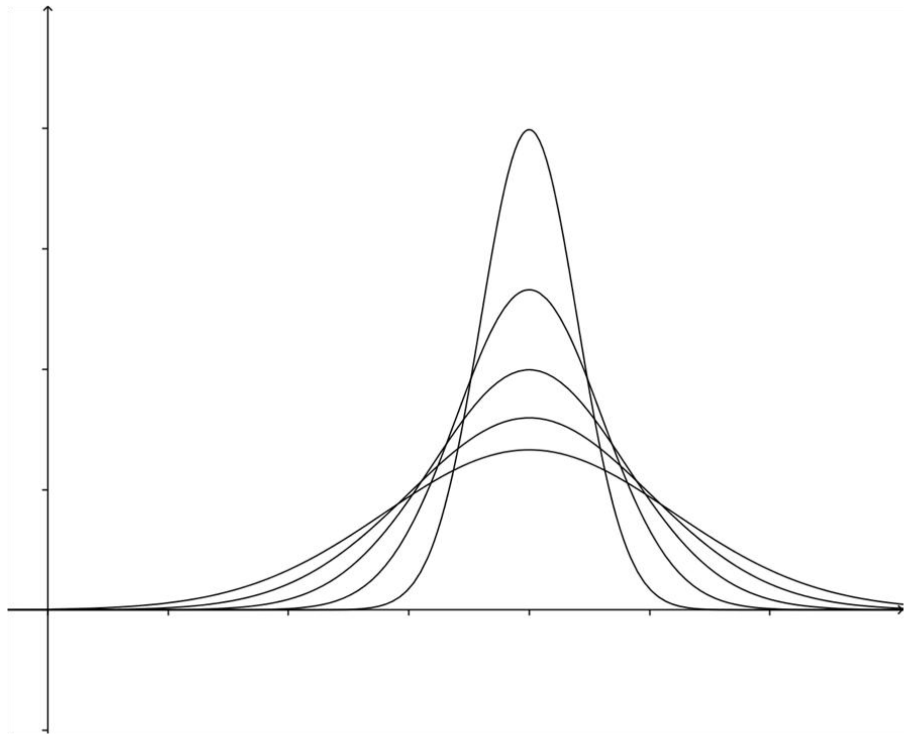
Construction de la fonction de densité d'une distribution normale

La loi normale est une distribution statistique dont l'histogramme est idéalement enveloppé par une courbe en forme de cloche. Les tailles d'un groupe d'individus, les pointures de chaussures, par exemple, sont distribuées ainsi. Sur le graphique, nous présentons différentes distributions normales de même moyenne et d'écarts type différents.



Il convient alors de définir une fonction mathématique dont la représentation graphique a l'allure ci-dessus. Voyons en le principe.

1) <u>Préliminaire</u>

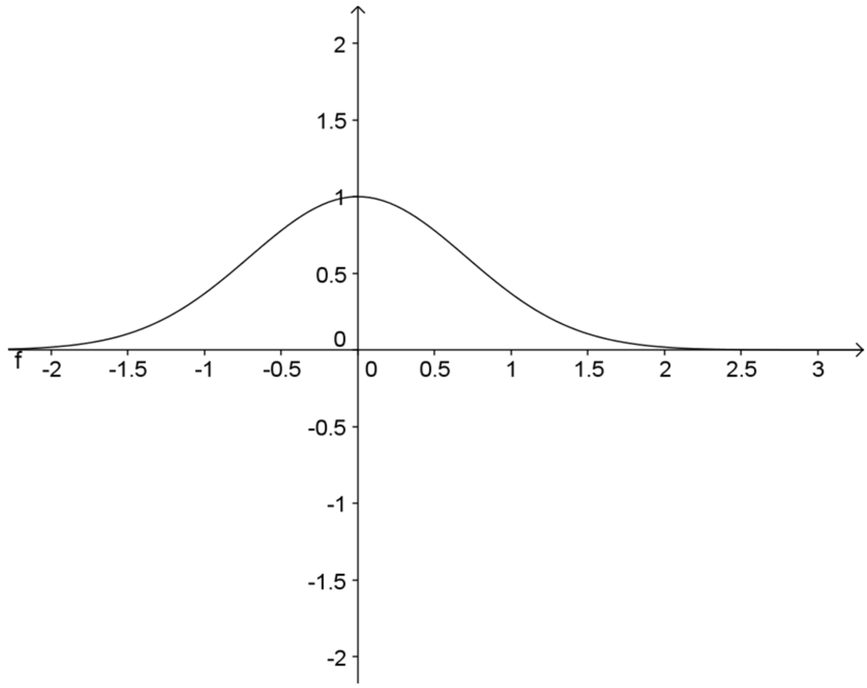
Soit f une fonction réelle définie sur $[0;+\infty[$. Considérons la fonction g définie sur $]-\infty;0]$ par :

$$g(x) = f(-x)$$

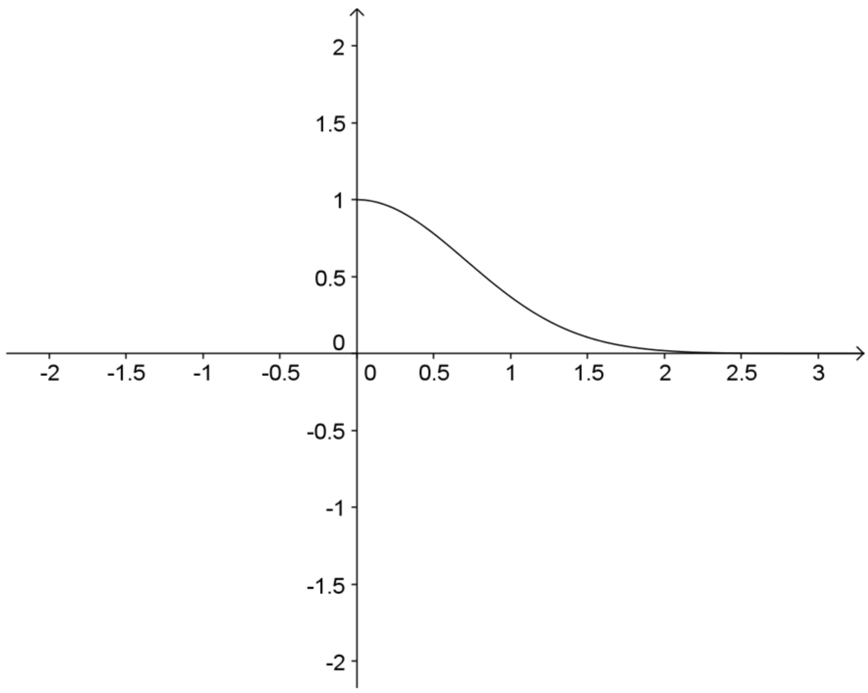
Alors la courbe représentative de g est la symétrique de celle de f par rapport à l'axe $(O\ y)$.

2) <u>Fonction de référence</u>

Commençons par chercher une fonction ayant l'allure précédente, mais telle que sa valeur maximale soit 1 et que $(O\ y)$ soit axe de symétrie.



Pour ce faire cherchons en l'expression d'abord sur $[0;+\infty[$, donc correspondant à un graphique de la sorte :



Une telle fonction a donc pour caractéristique de décroître et de valoir 1 en 0, ce qui rappelle l'exponentielle e^{-x}. Ainsi en définissant la partie symétrique, nous aurions :

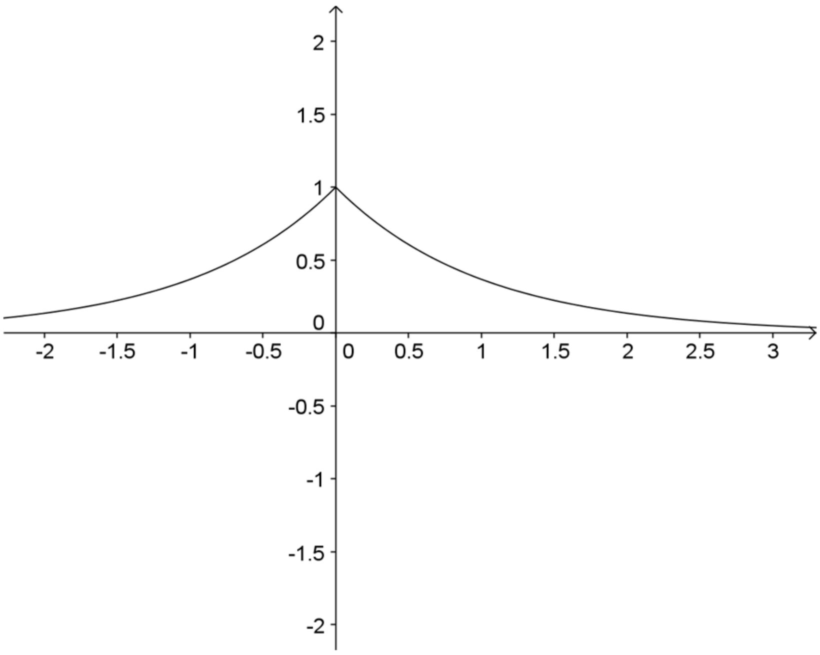
$$f(x) = e^{-x} \text{ sur } [0; +\infty[$$

$$f(x) = e^{x} \text{ sur }]-\infty; 0[$$

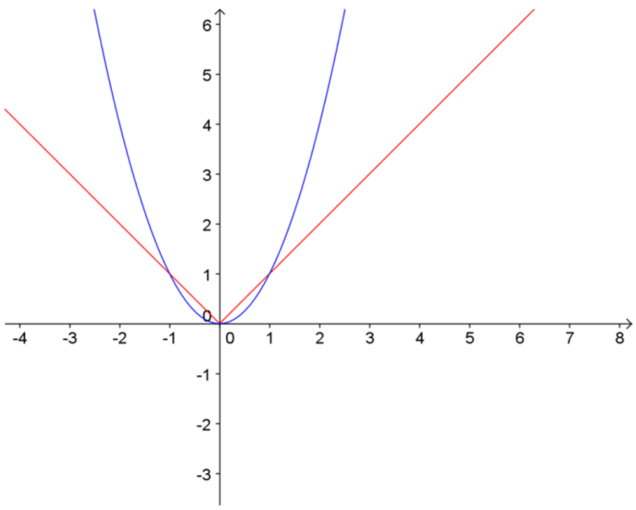
Ce qui peut s'écrire de façon plus compacte sur $\mathbb{R}$:

$$f(x) = e^{-|x|}$$

La courbe a alors l'allure suivante :



Mais elle présente un point anguleux en 0 en raison d'une dérivée à gauche et d'une dérivée à droite différentes. Un phénomène analogue se produit avec la fonction valeur absolue mais on constate qu'en prenant la fonction carré, on obtient une fonction de mêmes variations et sans point anguleux



En appliquant la même recette, nous sommes amenés à considérer la fonction définie sur $\mathbb{R}$ par :

$$f(x) = e^{-x^2}$$

3) Fonction déduite

L'étude des variations de la fonction précédente montre que le graphique de cette fonction a l'allure en cloche souhaitée. Afin d'en faire une fonction de distribution (appelée densité) il ne nous reste plus qu'à en transformer légèrement l'expression de la sorte :

$$f(x) = A\, e^{-a\,x^2}$$

où A et a sont deux constantes strictement positives que nous déterminerons de telle sorte que f représente bien la distribution d'une variable aléatoire X centrée d'écart type σ. Nous devons ainsi avoir :

$$\begin{cases} \displaystyle\int_{-\infty}^{+\infty} f(x)\, dx = 1 \\ E(X) = \displaystyle\int_{-\infty}^{+\infty} x\, f(x)\, dx = 0 \\ V(X) = \displaystyle\int_{-\infty}^{+\infty} x^2\, f(x)\, dx - \big(E(X)\big)^2 = \sigma^2 \end{cases}$$

Notons que la seconde condition est réalisée car la fonction $x\, f(x)$ est impaire. Le système de condition devient alors :

$$\begin{cases} A \displaystyle\int_{-\infty}^{+\infty} e^{-a\,x^2}\, dx = 1 \\ A \displaystyle\int_{-\infty}^{+\infty} x^2\, e^{-a\,x^2}\, dx = \sigma^2 \end{cases}$$

Faisons le changement de variable suivant :

$$\begin{cases} \sqrt{a}\, x = t \\ \sqrt{a}\, dx = dt \end{cases}$$

Le système devient :

$$\begin{cases} \dfrac{A}{\sqrt{a}} \displaystyle\int_{-\infty}^{+\infty} e^{-t^2}\, dt = 1 \\ \dfrac{A}{a\sqrt{a}} \displaystyle\int_{-\infty}^{+\infty} t^2\, e^{-t^2}\, dt = \sigma^2 \end{cases}$$

Or en intégrant par partie ($u = t, v' = t\, e^{-t^2}, u' = 1, v = -0,5\ e^{-t^2}$:

$$\int_{-\infty}^{+\infty} t^2\, e^{-t^2}\, dt = \int_{-\infty}^{+\infty} t\left(t\, e^{-t^2}\right) dt$$

$$= \left[-0,5\ t\, e^{-t^2}\right]_{-\infty}^{+\infty} + 0,5 \int_{-\infty}^{+\infty} e^{-t^2}\, dt$$

$$= 0,5 \int_{-\infty}^{+\infty} e^{-t^2}\, dt$$

Le système est alors équivalent au suivant

$$\begin{cases} A \displaystyle\int_{-\infty}^{+\infty} e^{-t^2}\, dt = \sqrt{a} \\ \dfrac{0,5}{a} = \sigma^2 \end{cases}$$

Finalement :

$$\begin{cases} A = \dfrac{\sqrt{a}}{\int_{-\infty}^{+\infty} e^{-t^2}\, dt} \\ a = \dfrac{1}{2\,\sigma^2} \end{cases}$$

La définition de f fait donc intervenir le calcul de l'intégrale impropre :

$$\int_{-\infty}^{+\infty} e^{-t^2}\, dt$$

4) Calcul de l'intégrale impropre précédente

L'idée est de calculer cette intégrale sur $[0;+\infty[$ en épousant le plus loin possible la fonction par une fonction majorante et une fonction minorante d'expressions plus simples (polynomiale ou rationnelle).

Pour ce faire on note que la fonction exponentielle e^{-t} peut être encadrée de la sorte sur $]-1;+\infty[$:

$$\boxed{1-t \leq e^{-t} \leq \frac{1}{1+t}}$$

Notons que l'inégalité supérieure équivaut à celle-ci :

$$1+t \leq e^{t}$$

Il suffit donc pour prouver l'encadrement de montrer que l'on a sur $\mathbb{R}$:

$$1+t \leq e^{t}$$

Ceci est aisé à car e^{t} est une fonction convexe et $y=1+t$ est sa tangente en 0. L'inégalité traduit que la courbe est au dessus de sa tangente en 0.

On peut retrouver ce résultat facilement en étudiant les variations de $g(t)=e^{t}-1-t$

Maintenant que l'encadrement est établi, l'idée est de considérer que cet encadrement colle bien à la fonction au voisinage de 0. Si on se donne alors un intervalle $[0;a]$, en prenant a suffisamment grand, pour tous les nombres t de $[0;a]$ et tous les entiers naturel n suffisamment grands , nous devrions avoir l'encadrement suivant très proche d'égalités :

$$1-\frac{t}{n} \leq e^{-\frac{t}{n}} \leq \frac{1}{1+\frac{t}{n}}$$

Voilà l'idée de départ. Précisons là en toute rigueur. Nous avons :

$$\forall\, n \in \mathbb{N}\ \ \forall\, t \in\ [0;n]\ :0 \leq 1-\frac{t}{n} \leq e^{-\frac{t}{n}} \leq \frac{1}{1+\frac{t}{n}}$$

En élevant cet encadrement à la puissance n, il en découle :

$$\boxed{\forall\, n \in \mathbb{N}\ \ \forall\, t \in\ [0;n]\ :\left(1-\frac{t}{n}\right)^{n} \leq e^{-t} \leq \left(1+\frac{t}{n}\right)^{-n}}$$

Soit en posant $t = x^2$

$$\forall\, n \in \mathbb{N} \;\; \forall\, x \in \left[0; \sqrt{n}\right] \; : \left(1 - \frac{x^2}{n}\right)^n \leq e^{-x^2} \leq \left(1 + \frac{x^2}{n}\right)^{-n}$$

D'où par comparaison :

$$\int_0^n \left(1 - \frac{x^2}{n}\right)^n dx \leq \int_0^n e^{-x^2} dx \leq \int_0^n \left(1 + \frac{x^2}{n}\right)^{-n} dx$$

Examinons les deux intégrales minorante et majorante. La minorante suggère un changement de variable tel que :

$$1 - \frac{x^2}{n} = cos^2(t)$$

Soit :

$$\begin{cases} x = \sqrt{n}\, sin(t) \\ dx = \sqrt{n}\, cos(t) dt \end{cases}$$

avec les changements de borne :

$$\begin{cases} x = 0 \;\to t = 0 \\ x = n \;\to t = \dfrac{\pi}{2} \end{cases}$$

La majorante suggère un changement de variable tel que :

$$1 + \frac{x^2}{n} = 1 + tan^2(t) = \frac{1}{cos^2(t)}$$

Soit :

$$\begin{cases} x = \sqrt{n}\, tan(t) \\ dx = \dfrac{\sqrt{n}}{cos^2(t)} dt \end{cases}$$

avec les changements de borne :

$$\begin{cases} x = 0 \;\to t = 0 \\ x = n \;\to t = \dfrac{\pi}{4} \end{cases}$$

Ainsi :

$$\boxed{\sqrt{n}\int_0^{\frac{\pi}{2}} cos^{2\,n+1}(t)dt \leq \int_0^{n} e^{-t^2}dt \leq \int_0^{\frac{\pi}{4}} cos^{2\,n-2}(t)dt \leq \sqrt{n}\int_0^{\frac{\pi}{2}} cos^{2\,n-2}(t)\,dt}$$

Cet encadrement fait intervenir deux suites extraites de la suite :

$$U_n = \int_0^{\frac{\pi}{2}} cos^n(t)\,dt$$

qui est une suite d'intégrales dites de Wallis que nous allons étudier afin d'en déterminer une expression explicite fonction de n.

5) <u>Intégrales de Wallis</u>

Ce sont les intégrales définies par les suites :

$$U_n = \int_0^{\frac{\pi}{2}} cos^n(t)\,dt$$

$$V_n = \int_0^{\frac{\pi}{2}} sin^n(t)\,dt$$

Mais en faisant le changement de variables suivant dans la première :

$$\begin{cases} t = \dfrac{\pi}{2} - x \\ dt = \ -dx \end{cases}$$

on note que :

$$U_n = V_n$$

Il suffit donc de s'intéresser à U_n. Or :

$$U_{n+1} - U_n = \int_0^{\frac{\pi}{2}} \left(cos^{n+1}(t) - cos^n(t)\right)dt = \int_0^{\frac{\pi}{2}} cos^n(t)\,(cos(t) - 1)\,dt$$

et :

$$\forall\ t \in \left]0; \frac{\pi}{2}\right[\ :\ cos^n(t)\ (cos(t) - 1) < 0$$

La fonction à intégrer étant continue, positive et non identiquement nulle sur l'intervalle d'intégration, on en déduit :

$$U_{n+1} - U_n < 0$$

Donc la suite U est strictement décroissante et d'autre part positive. La forme puissance de la fonction à intégrer invite à chercher une relation de récurrence pour U_n en l'écrivant :

$$U_n = \int_0^{\frac{\pi}{2}} cos(t)\ cos^{n-1}(t)\ dt$$

et en procédant à une intégration par partie en posant :

$$u' = cos(t), v = cos^{n-1}(t)$$

$$u = sin(t), v' = -(n-1)\ sint\ cos^{n-2}(t)$$

Afin que le procédé ait un sens il nous faut le considérer pour $n \geq 2$. On a alors :

$$U_n = [sin(t)\ cos^{n-1}(t)]_0^{\frac{\pi}{2}} + (n-1) \int_0^{\frac{\pi}{2}} sin^2(t)\ cos^{n-2}(t)\ dt$$

$$= (n-1) \int_0^{\frac{\pi}{2}} \left(1 - cos^2(t)\right) cos^{n-2}(t)\ dt$$

$$= (n-1) \left(\int_0^{\frac{\pi}{2}} cos^{n-2}(t)\ dt - \int_0^{\frac{\pi}{2}} cos^n(t)\ dt \right)$$

Finalement :

$$U_n = (n-1)(U_{n-2} - U_n)$$

$$U_n = (n-1)U_{n-2} - (n-1)U_{n-2}$$

$$n\ U_n = (n-1)U_{n-2}$$

$$\boxed{U_n = \frac{(n-1)}{n} U_{n-2}}$$

La suite U suit donc une relation de récurrence d'ordre 2 et il faut donc en déterminer les deux premiers termes.

$$U_0 = \int_0^{\frac{\pi}{2}} cos^0(t)\, dt = \int_0^{\frac{\pi}{2}} dt = \frac{\pi}{2}$$

$$U_1 = \int_0^{\frac{\pi}{2}} cos^1(t)\, dt = [\, sin(t)]_0^{\frac{\pi}{2}} = 1$$

Nous pouvons alors déterminer les termes de rang pair et ceux de rang impair :

$$U_{2\,n} = \frac{(2\,n-1)}{2\,n} U_{2\,n-2}$$

$$= \frac{(2\,n-1)}{2\,n}\frac{(2\,n-3)}{2\,n-2} U_{2\,n-4}$$

$$= \frac{(2\,n-1)}{2\,n}\frac{(2\,n-3)}{2\,n-2} \dots \frac{1}{2} U_0$$

$$= \frac{(2\,n)\,!}{\left[(2\,n)\left(2\,(n-1\,)\right)\dots(2\,\times 1)\right]^2}\frac{\pi}{2}$$

$$= \frac{(2\,n)\,!}{[2^n\, n!\,]^2}\frac{\pi}{2}$$

Finalement :

$$\boxed{U_{2\,n} = \frac{(2\,n)\,!}{2^{2\,n}\;(n!)^2}\frac{\pi}{2}}$$

De même :

$$U_{2\,n+1} = \frac{2\,n}{2\,n+1} U_{2\,n-1}$$

$$= \frac{(2\,n)}{(2\,n+1)}\frac{(2\,n-2)}{(2\,n-1)} U_{2\,n-3}$$

$$= \frac{(2\,n)}{(2\,n+1)}\frac{(2\,n-2)}{(2\,n-1)} \dots \frac{2}{3} U_1$$

$$=$$

Finalement :

$$\boxed{U_{2\,n+1} = \frac{2^{2\,n}\ (n!)^2}{(2\,n+1)\,!}}$$

Les formes obtenues invitent à effectuer le produit de deus termes consécutifs :

$$U_{2\,n}\,U_{2\,n+1} = \frac{(2\,n)\,!}{2^{2\,n}\ (n!)^2}\frac{\pi}{2}\,\frac{2^{2\,n}\ (n!)^2}{(2\,n+1)\,!}$$

Soit :

$$\boxed{U_{2\,n}\,U_{2\,n+1} = \frac{\pi}{2\,(2\,n+1)}}$$

Or :

$$\forall\,n \in \mathbb{N} :\ U_{2\,n+2}\ \leq\ U_{2\,n+1}\ \leq U_{2\,n}$$

Donc :

$$\frac{2\,n+1}{2\,n+2}U_{2\,n}\ \leq\ U_{2\,n+1}\ \leq U_{2\,n}$$

$$\frac{2\,n+1}{2\,n+2}\leq\ \frac{U_{2\,n+1}}{U_{2\,n}}\ \leq 1$$

On en déduit par théorème des gendarmes :

$$\lim_{n\to+\infty}\frac{U_{2\,n+1}}{U_{2\,n}} = 1$$

Soit :

$$U_{2\,n+1} \sim U_{2\,n}$$

Donc

$${U_{2\,n}}^2 \sim \frac{\pi}{2\,(2\,n+1)}$$

$$\boxed{U_{2\,n} \sim U_{2\,n+1} \sim \frac{\sqrt{\pi}}{2\,\sqrt{n}}}$$

Or

$$U_{2\,n} = \frac{(2\,n-1)}{2\,n}U_{2\,n-2}$$

Donc :

$$U_{2n-2} \sim U_{2n} \sim \frac{\sqrt{\pi}}{2\sqrt{n}}$$

6) Valeur de l'intégrale impropre

Reprenons l'encadrement :

$$\sqrt{n}\, U_{2n+1} \leq \int_0^n e^{-t^2} dt \leq \sqrt{n}\, U_{2n-2}$$

Nous avons :

$$\sqrt{n}\, U_{2n+1} \sim \sqrt{n}\ \frac{\sqrt{\pi}}{2\sqrt{n}} \sim \frac{\sqrt{\pi}}{2}$$

$$\sqrt{n}\, U_{2n-2} \sim \sqrt{n}\ \frac{\sqrt{\pi}}{2\sqrt{n}} \sim \frac{\sqrt{\pi}}{2}$$

On en déduit par théorème des gendarmes :

$$\int_0^n e^{-t^2} dt \sim \frac{\sqrt{\pi}}{2}$$

Donc, l'intégrale impropre converge et :

$$\int_0^{+\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$$

D'où :

$$\boxed{\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}}$$

7) Construction de la fonction associée à une distribution normale

Les valeurs A et a définies précédemment sont donc :

$$\begin{cases} A = \dfrac{\sqrt{a}}{\sqrt{\pi}} = \dfrac{1}{\sigma\sqrt{2\,\pi}} \\ a = \dfrac{1}{2\,\sigma^2} \end{cases}$$

et la fonction associée à une distribution centrée d'écart type σ est :

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\,\pi}}\; e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}}$$

et la fonction associée à une distribution d'espérance m d'écart type σ s'obtient par translation :

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

7) Propriétés de la loi normale

Si X est une variable distribuée selon une loi normale d'espérance m et d'écart type σ, alors la variable $Y = (X - m)/\sigma$ est distribuée selon une loi normale d'espérance nulle et d'écart type unité appelée loi normale centrée réduite.

Preuve :

Soit $y \in \mathbb{R}$ alors :

$$P(Y < y) = P\left(\frac{X-m}{\sigma} < y\right)$$

$$= P(X < m + \sigma\, y)$$

$$= \int_{-\infty}^{m+\sigma\, y} \frac{1}{\sigma\sqrt{2\,\pi}}\; e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

Faisons le changement de variable :

$$\begin{cases} \dfrac{x-m}{\sigma} = t \\ dx = \sigma\, dt \end{cases}$$

avec les changements de borne :

$$\begin{cases} x = m + \sigma\, y \ \rightarrow t = y \\ x = -\infty \ \rightarrow t = -\infty \end{cases}$$

alors :

$$P(Y < y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\,\pi}}\ e^{-\frac{1}{2}t^2} dt$$

et la densité associée est la dérivée, à savoir :

$$f_Y(y) = \frac{1}{\sqrt{2\,\pi}}\ e^{-\frac{1}{2}y^2}$$

ce qui correspond bien à une distribution normale d'espérance nulle et d'écart type égal à 1.

Les propriétés de la variance et de l'espérance permettent alors de vérifier :

$$E(X) = E(m + \sigma\, Y) = m + E(\sigma\, Y) = m + \sigma\, E(Y) = m$$

$$V(X) = V(m + \sigma\, Y) = V(\sigma\, Y) = \sigma^2\, V(Y) = \sigma^2$$

8) Autre méthode pour calculer l'intégrale impropre

Une méthode plus rapide permet de déterminer l'intégrale impropre en considérant l'intégrale double suivante :

$$A = \iint_{\mathbb{R}^2} e^{-\,(x^2+y^2)} dx\, dy$$

et en la calculant de deux façons :

1ère façon

L'intégrale double étant un volume situé entre la surface d'équation $z = e^{-(x^2+y^2)}$ et le plan $(O\ x\ y)$ on découpe ce volume en tranches de points d'abscisse x et d'épaisseur dx et on intègre ces volumes élémentaires sur y de $-\infty$ à $+\infty$

$$A = \int_{y=-\infty}^{y=+\infty} \left(\int_{x=-\infty}^{x=+\infty} e^{-(x^2+y^2)} dx \right) dy$$

$$\int_{y=-\infty}^{y=+\infty} e^{-y^2} \left(\int_{x=-\infty}^{x=+\infty} e^{-x^2} dx \right) dy$$

$$= \int_{x=-\infty}^{x=+\infty} e^{-x^2} dx \int_{y=-\infty}^{y=+\infty} e^{-y^2}\, dy$$

$$= \left(\int_{x=-\infty}^{x=+\infty} e^{-x^2} dx \right)^2$$

2ème façon :

on découpe ce volume en cylindres d'axe $(O\ z)$, de base annulaire de rayon intérieur r et de rayon extérieur $r + dr$ et on intègre ces volumes élémentaires sur r de 0 à $+\infty$

$$A = \int_{r=0}^{r=+\infty} e^{-r^2} d(\pi\, r^2)$$

$$= \int_{r=0}^{r=+\infty} 2\,\pi\, r\, e^{-r^2} dr$$

$$= \left[\pi\, e^{-r^2}\right]_0^{+\infty} = \pi$$

En comparant les deux calculs de A on en déduit :

$$\int_{x=-\infty}^{x=+\infty} e^{-x^2} dx = \sqrt{\pi}$$