

La loi normale (ou loi de Laplace-Gauss ou loi de Gauss)

1 – Les fondamentaux

La distribution normale est une distribution théorique, en ce sens qu'elle est une idéalisation mathématique qui ne se rencontre jamais *exactement* dans la nature. Mais de nombreuses distributions réellement observées s'en rapprochent et ont cette fameuse forme de « cloche » (beaucoup d'individus autour de la moyenne, de moins en moins au fur à mesure qu'on s'en éloigne, et ceci de façon symétrique).

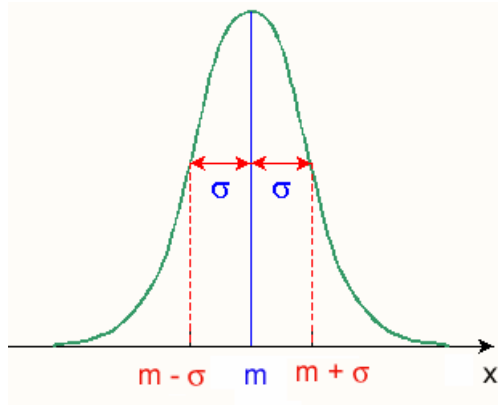
D'autre part, elle est très utilisée en statistiques inférentielles : nous verrons en particulier qu'une moyenne calculée sur un échantillon est une v.a. qui tend à suivre une loi normale quand la taille de l'échantillon augmente, même si la population initiale a une tout autre distribution.

a – Sa forme : la courbe en cloche

La loi normale de paramètres m et σ , notée $N(m, \sigma)$, est définie sur \mathbb{R} par la densité :

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

dont la représentation graphique est la suivante :



- Notons que :
- la droite $x = m$ est axe de symétrie
 - les points d'inflexion sont situés à une distance σ de cet axe de symétrie

b – Le théorème Central-limit

Le TCL sera très précieux puisqu'il nous explique que si on fait la somme d'un très grand nombre de variables aléatoires de loi quelconque, cette somme suit approximativement une loi normale (en fait, sans rentrer dans le détail des hypothèses, il nous dit que la variable $X = X_1 + X_2 + \dots + X_n$ tend à suivre une loi normale quand n tend vers l'infini).

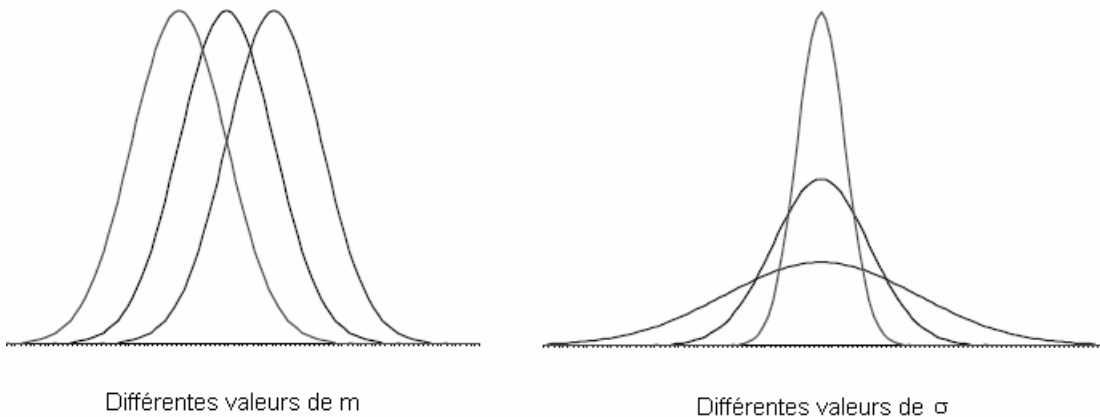
D'une part, cela nous permet de comprendre pourquoi autant de distributions observées dans la réalité ont approximativement cette forme de cloche : elles décrivent des phénomènes qui résultent de l'addition d'un grand nombre de causes de fluctuation indépendantes. Exemple : la taille d'un individu.

D'autre part, cela nous permettra d'approcher beaucoup de lois par une loi normale, pour peu que la variable étudiée s'exprime comme une somme d'un grand nombre de variables indépendantes. C'est le cas notamment de la variable binomiale (somme de n variables de Bernoulli indépendantes), dont la loi «tend à prendre la forme d'une cloche » quand n augmente. Cela reste possible même quand on ne connaît pas loi des variables X_i .

2 – Espérance et variance

Soit X une v.a. qui suit la loi $N(m,\sigma)$. Par raison de symétrie: $E(X) = m$ et on montre facilement que $V(X) = \sigma^2$, donc le paramètre σ correspond à l'écart-type (d'où les notations...)

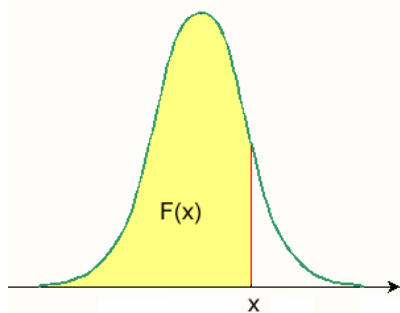
Ainsi grâce à ses 2 paramètres, la loi normale permet de décrire des distributions de moyenne quelconque (on translate la courbe vers la gauche ou vers la droite), et de dispersion quelconque (on rapproche ou on écarte le point d'inflexion)



3 – Calculs de probabilités sur une loi normale

a – Un gros inconvénient : on ne sait pas exprimer $F(x)$ en fonctions de x

On ne connaît pas de primitive de la fonction e^{-x^2} , donc on ne sait pas donner l'expression algébrique de la fonction de répartition $F(x)$.



Comment dans ces conditions calculer les probabilités de tomber entre telle ou telle valeur? Par des techniques de calcul numérique (en mesurant l'aire sous la courbe pour différentes valeurs de x), on a pu constituer des tables donnant $F(x)$. Ces tables figurent en annexe de la plupart des manuels de probabilités, et sont intégrées dans certains logiciels (fonctions d'Excel par exemple). Pour tous les calculs, on se ramène à la fonction de répartition de la loi $N(0,1)$, dite loi normale centrée-réduite.

b – La loi normale centrée-réduite

Centrer et réduire une variable, c'est raisonner en nombre d'écarts-types par rapport à la moyenne. Par ex., si le poids d'un foie gras est une variable de moyenne 550 g et d'écart-type 100 g, on dira d'un foie de canard de $x=650$ g que, en donnée centrée-réduite, il pèse $t=1$ (sous-entendu : 1 écart-type de plus que la moyenne), alors que le poids centré-réduit d'un foie de $x'=500$ g sera de $t'=-0,5$ (un demi-écart-type en dessous de la moyenne).

Tous les événements relatifs à X peuvent être aussi bien exprimés en fonction de T . Ainsi, il est équivalent de dire :

« le poids d'un foie est compris entre 500g et 650g » : $(500 < X < 660)$ *ou*
 « le poids centré-réduit est compris entre -0,5 et 1 » : $(-0,5 < T < 1)$

avec T égal ici à : $\frac{X - 550}{100}$, et plus généralement : $T = \frac{X - m}{\sigma}$

La variable centrée-réduite T a pour **espérance 0** et pour **écart-type 1** car :

$$E(T) = E\left(\frac{X-m}{\sigma}\right) = \frac{1}{\sigma}(E(X)-m) = 0 \quad \text{puisque} \quad E(X)=m$$

$$V(T) = V\left(\frac{X-m}{\sigma}\right) = \frac{1}{\sigma^2}(V(X)) = 1 \quad \text{puisque} \quad V(X)=\sigma^2$$

Donc la densité de probabilité de la loi normale centrée-réduite N(0,1) s'écrit :

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

Plutôt que f et F, on note généralement y la densité, et Π la fonction de répartition de la loi N(0,1). La table donnant, pour différentes valeurs de t, les valeurs de $\Pi(t)$, soit $P(T \leq t)$ est jointe en annexe.

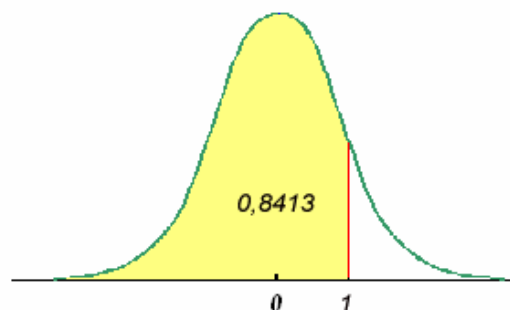
On y lit par exemple :

$$\Pi(0) = 0,5$$

$$\Pi(1) = 0,8413$$

$$\Pi(0,5) = 0,6915$$

$$\Pi(1,96) = 0,9750$$

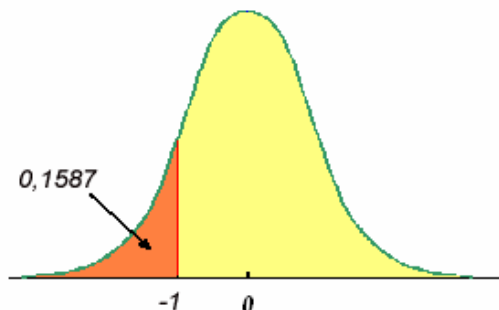


dont on déduit par symétrie :

$$\Pi(-1) = 1 - \Pi(1) = 1 - 0,8413 = 0,1587$$

$$\Pi(-0,5) = 1 - \Pi(0,5) = 1 - 0,6915 = 0,3085$$

$$\Pi(-1,96) = 1 - \Pi(1,96) = 1 - 0,9750 = 0,0250$$



c – Exemples de calcul sur une loi normale

La v.a. X, poids d'un foie gras, suit une loi N(550 ;100). Quelle est la probabilité pour qu'un foie gras pèse moins de 650g, plus de 746g, moins de 500g, entre 550 et 600g ?

$$P(X < 650) = P\left(T < \frac{650-550}{100}\right) = P(T < 1) = \Pi(1) = 84,13\%$$

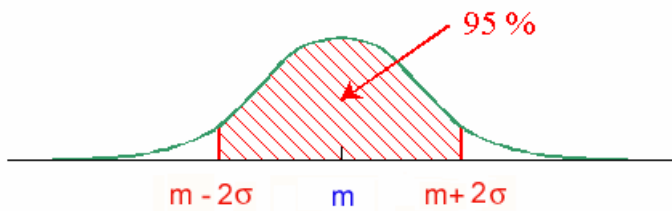
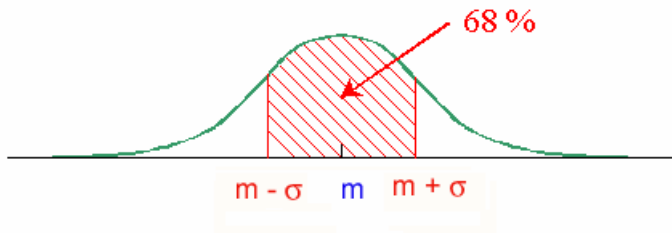
$$P(X > 746) = P\left(T > \frac{746-550}{100}\right) = P(T > 1,96) = 1 - P(T \leq 1,96) = 1 - \Pi(1,96) = 1 - 0,9750 = 2,5\%$$

$$P(X < 500) = P\left(T < \frac{500-550}{100}\right) = P(T < -0,5) = \Pi(-0,5) = 1 - \Pi(0,5) = 1 - 0,6915 = 30,85\%$$

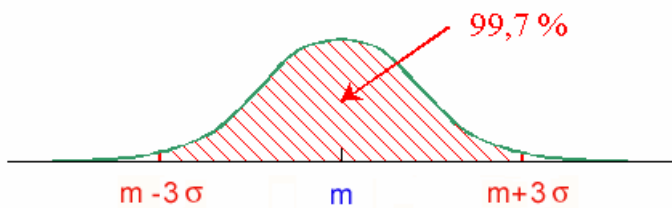
$$P(550 < X < 600) = P(0 < T < 0,5) = \Pi(0,5) - \Pi(0) = 0,8413 - 0,5 = 34,13\%$$

Rappelons que pour une variable continue, il n'y a pas de différence entre $P(X < k)$ et $P(X \leq k)$ car la probabilité attachée à la valeur k est nulle.

d- Quelques ordres de grandeur utiles à retenir



Une variable normale a « 95 chances sur 100 d'être située entre : moyenne moins 2 écarts-types et moyenne plus 2 écarts-types » (la vraie valeur n'est pas 2 mais 1,96)



Une variable normale est presque certainement située entre : moyenne moins 3 écarts-types et moyenne plus 3 écarts-types

3 – Stabilité de la loi normale

Une combinaison linéaire de variables normales indépendantes est elle-même une variable normale.

Ainsi si $X_1 \leftrightarrow N(m_1, \sigma_1)$ et $X_2 \leftrightarrow N(m_2, \sigma_2)$, X_1 et X_2 étant indépendantes
si a_1 et a_2 sont 2 réels

alors : $X = a_1 X_1 + a_2 X_2$ suit également une loi normale (dont les paramètres peuvent être calculés en utilisant les propriétés de l'espérance et de la variance).

Par exemple, $S = X_1 + X_2$ est une variable normale de paramètres :

$$E(S) = E(X_1 + X_2) = E(X_1) + E(X_2) = m_1 + m_2$$

$$V(S) = V(X_1 + X_2) = V(X_1) + V(X_2) \quad \text{car les variables sont indépendantes}$$

$$= \sigma_1^2 + \sigma_2^2$$

donc $\sigma_S = \sqrt{\sigma_1^2 + \sigma_2^2}$ et finalement : $S \leftrightarrow N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

4 – Approximation de la loi binomiale et de la loi de Poisson par la loi normale

Nous avons vu que la loi binomiale $B(n, p)$ est d'autant plus symétrique que p est proche de 50% et qu'elle prend une forme en cloche quand n augmente (d'autant plus vite que p est proche que 0,5...)
Le TCL donne une justification à ce phénomène.

Une valeur de p très différente de 0,5 (p petit ou alors $q=1-p$ petit) pourra être compensée par une grande valeur de n et on accepte généralement de remplacer la loi binomiale par la loi normale

lorsque les produits np et nq sont supérieurs à 15 ou 20. Dans ce cas, on approchera la loi $B(n,p)$ par la loi normale de même espérance et de même écart-type, soit $N(np, \sqrt{npq})$

De même on pourra remplacer la loi de Poisson $P(m)$ par une loi normale $N(m, \sqrt{m})$ dès que m est supérieur à 15 ou 20.

Exemple : on estime que la probabilité pour qu'une graine ait perdu son pouvoir germinatif après 3 ans de conservation est de 70%. Sur un échantillon de 100 graines conservées depuis 3 ans quelle est la probabilité pour que moins de 25 germent ?

Notons p la probabilité qu'une graine germe : $p = 0,3$ et considérons que l'échantillon est indépendant. Notons X la v.a. « nombre de graines qui germent parmi les 100 ».

X suit la loi $B(100 ; 0,3)$ et on cherche : $P(X < 25)$ qui peut s'écrire aussi $P(X \leq 24) = p_0 + p_1 + \dots + p_{24}$ avec $p_k = \binom{100}{k} 0,3^k 0,7^{100-k}$

Le calcul exact est trop fastidieux pour être fait à la main. On peut alors :

- soit utiliser un logiciel, par exemple la fonction d'Excel=LOI.BINOMIALE(24;100;0,3;1) qui donne $P(X \leq 24) = 0,114$
- soit calculer une valeur approchée en remplaçant cette loi binômiale par une loi normale. C'est possible car les produits np et nq sont assez grands (resp. 30 et 70). Les paramètres de cette loi seront :
 - o $m = np = 30$
 - o $\sigma = \sqrt{npq} = \sqrt{100 * 0,3 * 0,7} = 4,5826$

La variable aléatoire discrète $X \leftrightarrow B(100 ; 0,3)$ sera alors remplacée par la variable continue :

$$X_c \leftrightarrow N(30 ; 4,5826)$$

Un problème se pose alors : faut-il calculer $P(X_c < 25)$ ou $P(X_c \leq 24)$? Pour une variable continue, ces valeurs ne sont pas identiques... La meilleure approximation sera obtenue en prenant la valeur intermédiaire 24,5. C'est ce qu'on appelle la « correction de continuité ». Voir la justification page suivante.

$$P(X < 25) = P(X \leq 24) \cong P(X_c \leq 24,5) = \Pi\left(\frac{24,5 - 30}{4,5826}\right) = \Pi(-1,20) = 1 - \Pi(1,20) = 1 - 0,885 = 0,115$$

On peut constater que ceci fournit une excellente approximation de la vraie valeur puisque l'erreur est de l'ordre du millième.

Bibliographie :

<http://neumann.hec.ca/~p240/c162096/documents/NotesLoiNormale1x1.pdf>

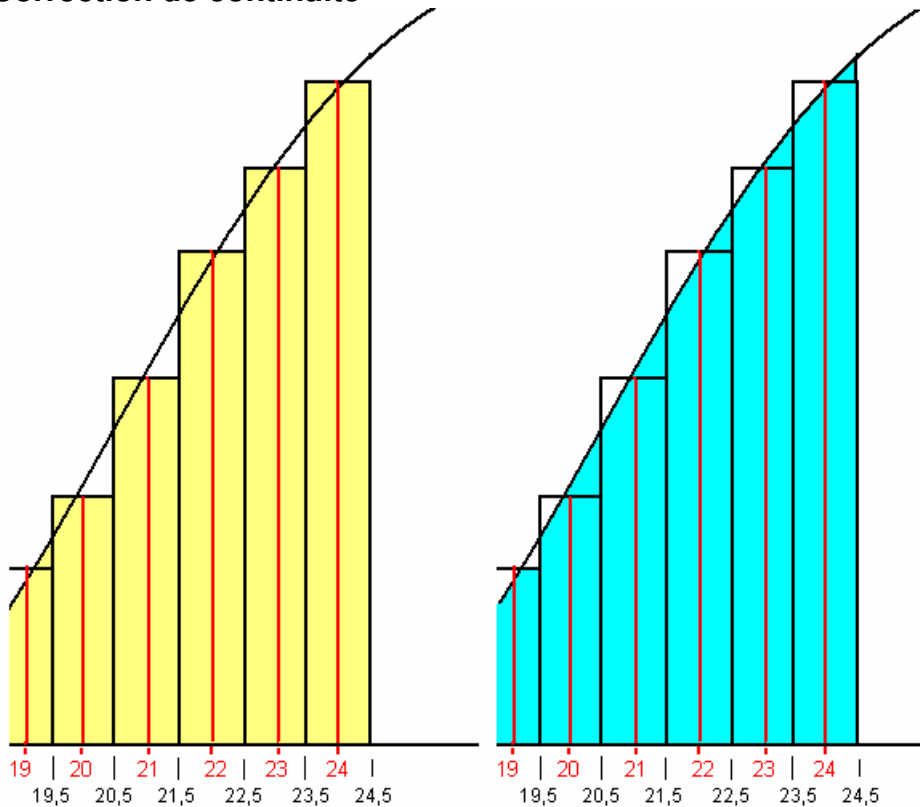
<http://www.astro.ulg.ac.be/cours/magain/stat/stat52.html>

<http://www.astro.ulg.ac.be/cours/magain/stat/stat51.html>

<http://www.up.univ-mrs.fr/veronis/cours/INFZ16/index.html?http://www.up.univ-mrs.fr/veronis/cours/INFZ16/ch6.html>

Annexes

Correction de continuité



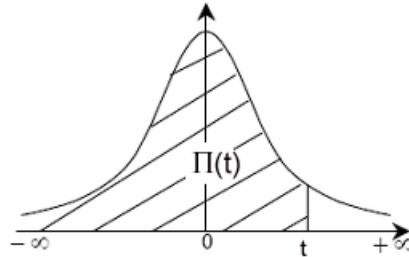
En jaune: la valeur exacte que l'on veut calculer. En effet $P(X \leq 24) = p_0 + p_1 + \dots + p_{24}$, ce qui correspond à la somme des hauteurs de bâtons rouges du diagramme en bâton de la loi binômiale. Cette somme est égale à la surface des rectangles jaunes puisque ces rectangles ont pour hauteur les p_i et pour base 1.

En bleu: ce qu'on calcule en prenant $P(X_c \leq 24,5)$, qui correspond à la surface sous la courbe de densité à gauche du point 24,5. On voit bien que l'approximation serait moins bonne en s'arrêtant à 24 ou en allant jusqu'à 25.

On pratique la correction de continuité chaque fois qu'on approche une loi discrète par une loi continue (en fait chaque fois qu'on hésite entre 2 valeurs comme entre 24 et 25 ici)

Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à t



$$\Pi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Table pour les grandes valeurs de t :

t	3	3,2	3,4	3,6	3,8	4	4,2	4,4	4,6	4,8
$\Pi(t)$	0,99865003	0,99931280	0,99966302	0,99984085	0,99992763	0,99996831	0,99998665	0,99999458	0,99999789	0,99999921