

CHAPITRE III : LA LOI NORMALE

I La loi normale

La loi normale ou loi de Laplace-Gauss est une des distributions que l'on rencontre le plus souvent en pratique. C'est, en effet, la loi qui s'applique à une variable statistique qui est la résultante d'un grand nombre de causes indépendantes, dont les effets s'additionnent et dont aucune n'est prépondérante.

Un peu d'histoire mathématique : La loi normale est, entre autres, la loi des erreurs de mesure d'un phénomène physique. Historiquement, c'est par l'étude du comportement de ces erreurs que Laplace (1777, 1812) puis Gauss (1821) ont mis en évidence l'expression mathématique de cette loi.

Mais tout le mérite revient à Moivre (1718), qui, en appliquant la formule de son ami Stirling, en a obtenu l'expression en cherchant des équivalents des probabilités binomiales.

Extrait du cours de l'UPPA : Professeur : Marc Artzrouni : la loi normale se rencontre souvent dans la nature et elle est une bonne approximation pour une variable qui est symétrique autour d'une valeur moyenne :

Tailles biologiques, niveau d'une rivière chaque jour, volume des échanges à la bourse, tension artérielle d'un individu...

Fin de l'extrait du cours de l'UPPA : Professeur : Marc Artzrouni

1) Définition

a) La loi de probabilité normale

Définition : La variable aléatoire **normale** X est une variable continue pouvant prendre n'importe quelle valeur entre $-\infty$ et $+\infty$ avec la densité de probabilité f définie par

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

où μ et σ sont deux paramètres réels tels que $\sigma > 0$.

La notation $dnorm(x, \mu, \sigma)$ est celle du logiciel MathCad.

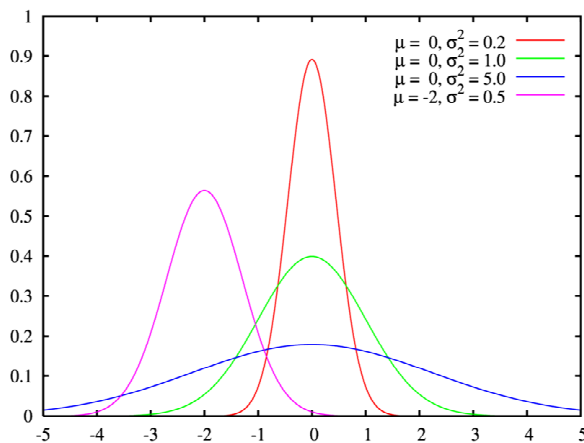
La fonction de répartition, qui représente la probabilité que la variable aléatoire X ait une valeur inférieure ou égale à x, a pour expression :

$$\text{pnorm}(x, \mu, \sigma) = F_X(x) = p(X \leq x) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^x e^{-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2} dy$$

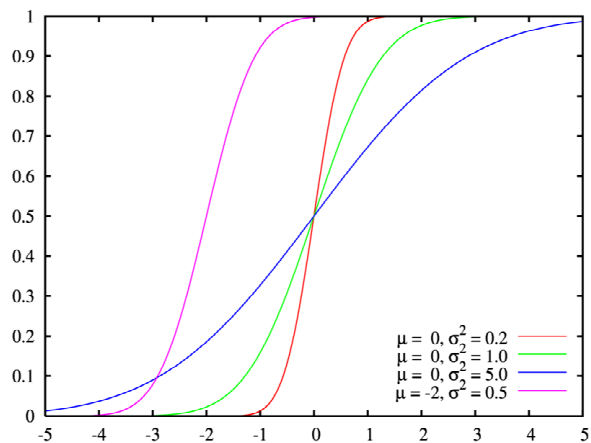
Symboliquement, nous noterons X par $N_{\mu, \sigma}$ pour indiquer que la variable aléatoire X suit une loi normale de paramètres μ et σ .

La notation $\text{pnorm}(x, \mu, \sigma)$ est celle du logiciel MathCad.

Représentation graphique de la fonction de la densité de probabilité



Représentation graphique de la fonction de la fonction de répartition



Légende de la courbe de gauche : la courbe dont le maximum est 0,9 correspond à $\mu = 0$ et $\sigma^2 = 0,2$, celle dont le maximum est 0,58 correspond à $\mu = -2$ et $\sigma^2 = 0,5$, celle dont le maximum est 0,4 correspond à $\mu = 0$ et $\sigma^2 = 1$ et la quatrième correspond à $\mu = 0$ et $\sigma^2 = 5$.

Légende de la courbe de droite : la courbe « commençant à $-1,2$ » correspond à $\mu = 0$ et $\sigma^2 = 0,2$, celle « commençant à -4 » correspond à $\mu = -2$ et $\sigma^2 = 0,5$, celle « commençant à -3 » correspond à $\mu = 0$ et $\sigma^2 = 1$ et la quatrième correspond à $\mu = 0$ et $\sigma^2 = 5$.

b) Loi de probabilité normale réduite

Définition : Soit $X = N_{\mu, \sigma}$ une variable aléatoire suivant une loi normale de paramètres μ et σ . On effectue le changement de variable

$$Z = \frac{X - \mu}{\sigma}$$

Alors Z est une variable aléatoire qui suit **une loi normale de paramètres 0 et 1**.

Z se note $N_{0,1}$ et est appelée **variable aléatoire normale centrée réduite**.

2) Caractéristiques de la loi normale

a) L'espérance mathématique

Définition : **L'espérance mathématique (ou moyenne)** de la loi normale $N_{\mu, \sigma}$ est

$$E(N_{\mu, \sigma}) = \mu$$

b) La variance et l'écart-type

Définition : **La variance** de la loi normale $N_{\mu, \sigma}$ est

$$V(N_{\mu, \sigma}) = \sigma^2$$

L'écart-type de la loi normale $N_{\mu, \sigma}$ est

$$\sigma(N_{\mu, \sigma}) = \sigma$$

Remarque : L'espérance mathématique (ou moyenne) de la loi normale centrée réduite $N_{0,1}$ est égale à **0** et son écart-type est égal à **1**.

3) Allure de la courbe de densité de la loi normale

La courbe de densité de probabilité de la loi de Laplace-Gauss se présente comme une courbe symétrique à un seul mode. Sa forme particulière lui a valu la dénomination de courbe en cloche.

Les observations sont groupées autour de la moyenne de la manière suivante :

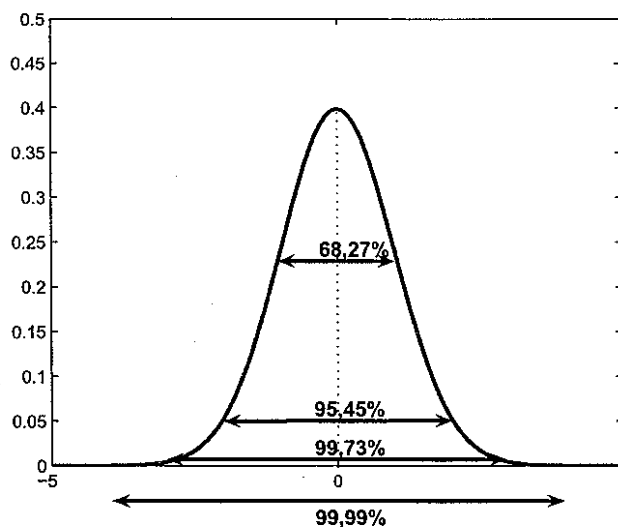
50 % des observations sont dans l'intervalle $\left[\mu - \frac{2}{3}\sigma ; \mu + \frac{2}{3}\sigma \right]$;

68,27 % des observations sont dans l'intervalle $[\mu - \sigma ; \mu + \sigma]$;

95,45 % des observations sont dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$;

99,73 % des observations sont dans l'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$;

99,99 % des observations sont dans l'intervalle $[\mu - 4\sigma ; \mu + 4\sigma]$.

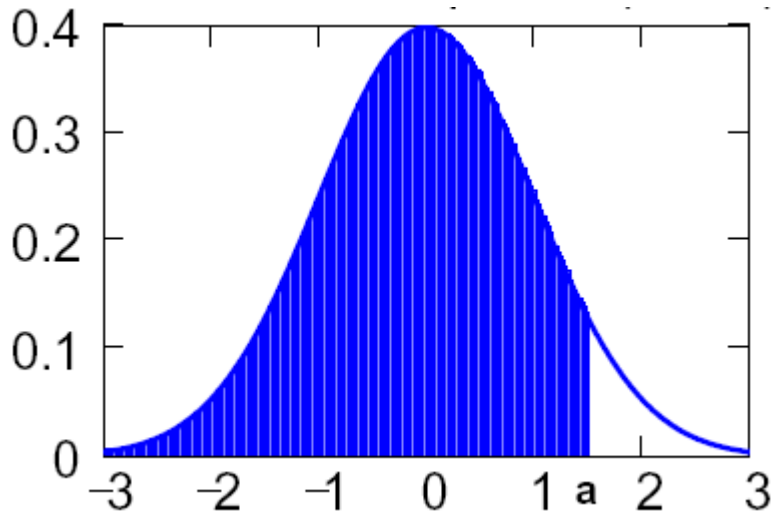


En pratique, la quasi-totalité des unités est donc rassemblée dans un intervalle de largeur huit écart-types autour de la moyenne.

4) Allure de la courbe de densité de la loi normale centrée réduite $N_{0,1}$

La fonction de répartition notée **pnorm**, de la loi normale centrée réduite est représentée par la courbe cumulative. Celle-ci est symétrique par rapport au point d'inflexion de coordonnées $(0, \frac{1}{2})$.

La valeur $\text{pnorm}(a, 0, 1) = p(N_{0,1} \leq a)$ de la fonction de répartition est égale à l'aire hachurée comprise entre la courbe de densité de probabilité, l'axe des abscisses et la droite d'équation $x = a$.



La courbe de densité de probabilité de $N_{0,1}$ est symétrique par rapport à la droite d'équation $x = 0$ donc

$$\mathbf{pnorm(-a, 0, 1) = 1 - pnorm(a, 0, 1)}$$

5) Conditions d'applications

a) Le théorème Central-Limit (ou théorème de la limite centrale)

La loi normale est d'application très générale.

Soient X_1, X_2, \dots, X_n une suite de n variables aléatoires indépendantes, de même loi, de même espérance mathématique notée μ , de même variance notée σ^2 . Alors la variable aléatoire $Z_n = \frac{(X_1 + X_2 + \dots + X_n) - n\mu}{\sigma\sqrt{n}}$ converge en loi vers une variable

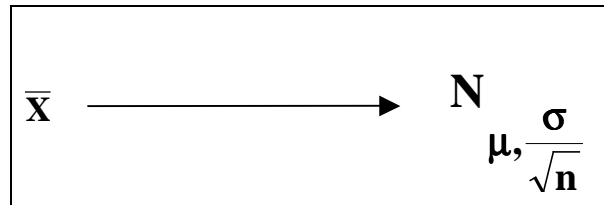
aléatoire normale centrée réduite c'est-à-dire $F_{Z_n}(z) = p(Z_n \leq z)$ tend vers

$\mathbf{pnorm}(z, 0, 1)$ lorsque n tend vers $+\infty$.

Par conséquent, les phénomènes qui peuvent être considérés comme engendrés par un grand nombre de causes élémentaires de fluctuations agissantes de façon indépendante, seront généralement susceptibles d'être représentés par la loi normale ; ainsi les caractéristiques (diamètre, poids...) d'une pièce fabriquée en série sont soumises à un grand nombre de causes de perturbations : variations de la température, différence d'homogénéité de la matière première...

b) Loi de la moyenne d'un gros échantillon

La loi de probabilité de la moyenne \bar{X} d'un échantillon de taille n , choisi avec remise dans une population de moyenne μ et d'écart-type σ , est approximativement, quelle que soit la loi de distribution X , une loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$

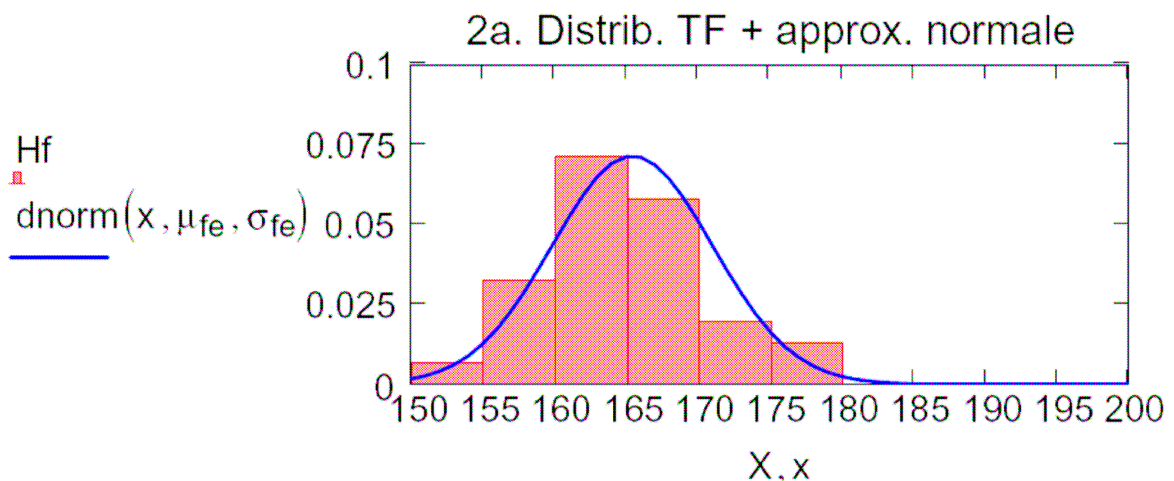


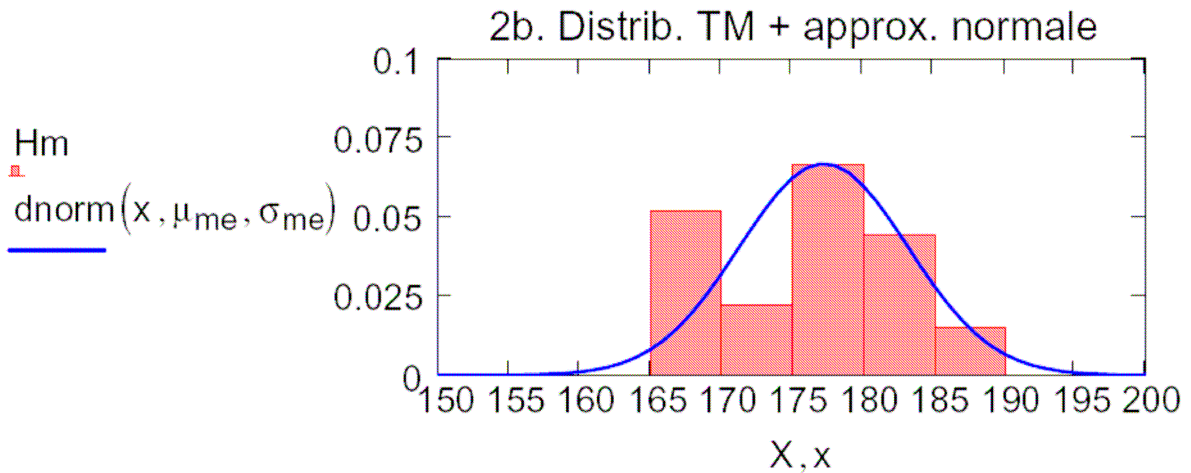
Proposition : Ce résultat est généralement considéré comme valable lorsque $n > 30$.

c) La loi normale comme modèle de la taille des étudiants : (Extrait du cours de l'UPPA : Professeur : Marc Artzrouni)

i) Approximation normale

Chaque taille de filles et de garçons d'un amphithéâtre peut être considérée comme le tirage d'une variable aléatoire notée TF ou TM (en anglais : F pour female et M pour male). Dans les graphiques ci-dessous sont représentées les fréquences observées pour les deux sexes, par groupe de taille quinquennal, et les densités normales qui les approximent. Les paramètres, μ_{fe} , σ_{fe} , μ_{me} , σ_{me} utilisés sont des estimateurs des espérances et écart-types théoriques mais inconnus μ_f , σ_f de TF et μ_m , σ_m de TM. (voir ci-dessous)





Les aires de chaque rectangle (ordonnée \times 5) donnent les fréquences correspondant à l'intervalle. Par exemple, $H_{m_6} = 0.0222$ et $H_{m_6} \times 5 = 0.111 = 11,1\%$ est la fréquence des hommes de taille entre 170 et 175 cm. De cette façon, la somme des aires est 1, tout comme l'aire totale sous les densités de probabilité $dnorm(\mu_{fe}, \sigma_{fe})$ et $dnorm(\mu_{me}, \sigma_{me})$. Les graphes montrent que ces densités sont des approximations des densités de probabilité des variables aléatoires TF et TM.

ii) Estimation des paramètres $\mu_f, \sigma_f, \mu_m, \sigma_m$

On estime μ_f , l'espérance, la valeur moyenne de TF, par l'unique réalisation que l'on

a de la variable aléatoire μ_{fe} définie par $\mu_{fe} = \frac{\sum TF_k}{n_f}$ (n_f = nombre total de filles

dans l'amphithéâtre) qui est la moyenne empirique ou expérimentale des tailles des femmes : $\mu_{fe} = 165.419$. De même, $\mu_{me} = 177.259$. (tout ce qui est dit dans ce paragraphe est identique pour les garçons.)

Cet estimateur est dit **sans biais** en ce sens que son espérance est égale au paramètre μ_f que l'on cherche à connaître.

$$\text{En effet, } E(\mu_{fe}) = E\left(\frac{\sum TF_k}{n_f}\right) = \frac{E\left(\sum TF_k\right)}{n_f} = \frac{\mu_f n_f}{n_f} = \mu_f$$

Pour σ_f , on rappelle la formule de Koenigs-Huygens : $\sigma_f^2 = E(TF^2) - \mu_f^2$.

L'estimateur naturel de σ_f^2 serait σ_{fe}^2 avec $\sigma_{fe}^2 = \frac{\sum (TF_k)^2}{n_f} - \mu_{fe}^2$ mais on montre que cet estimateur serait biaisé c'est-à-dire que l'on n'a pas $E(\sigma_{fe}^2) = \sigma_f^2$.

Pour σ_{fe}^2 , on va plutôt prendre la quantité voisine $\sigma_{fe}^2 = \frac{\sum (TF_k)^2}{n_f - 1} - \frac{n_f}{n_f - 1} \mu_{fe}^2$. On montre que cet estimateur σ_{fe}^2 , que l'on appelle **la variance empirique** est sans biais : son espérance est σ_f^2 , la quantité que l'on estime à l'aide de σ_{fe}^2 .

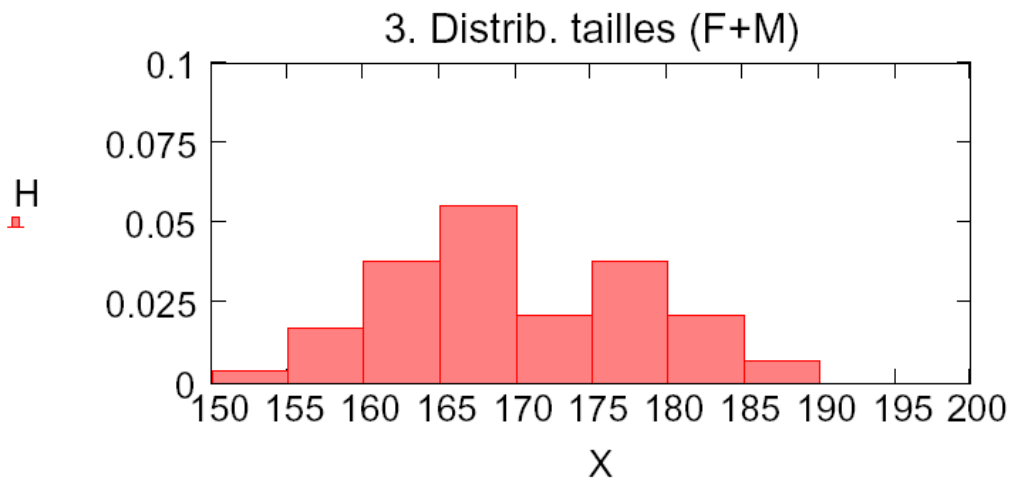
Application numérique :

Pour les calculs, on a besoin de : $n_f = 31$, $\sum_k TF_k = \text{somTF} = 5.128 \times 10^3$, la somme des carrés $\sum_k (TF_k)^2 = \text{som2TF} = 8.49208 \times 10^5$.

Calculer $\mu_{fe} = \frac{\text{somTF}}{n_f} =$

puis calculer $\sigma_{fe}^2 = \frac{\text{som2TF}}{n_f - 1} - \frac{n_f \left(\frac{\text{somTF}}{n_f} \right)^2}{n_f - 1} =$

Remarque : Si on avait pris les filles et les garçons ensemble, l'histogramme aurait eu l'allure suivante :



qui n'est pas très « normale » : distribution asymétrique avec les deux bosses qui correspondent aux maximums des deux distributions.

Fin de l'extrait du cours de l'UPPA : Professeur : Marc Artzrouni)

d) Somme de variables aléatoires indépendantes

La somme de deux variables normales indépendantes ayant respectivement pour paramètres (μ_1, σ_1) et (μ_2, σ_2) est elle-même une variable normale de moyenne

$$\mu = \mu_1 + \mu_2 \text{ et de variance } \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

6) Détermination pratique des probabilités : usage des tables de la loi normale

Par le changement de variable $T = \frac{X - \mu}{\sigma}$, si X suit la loi normale $N_{\mu, \sigma}$ alors T suit la

loi normale centrée réduite $N_{0,1}$

a) Table de la densité de probabilité dnorm : descriptif

Cette table donne la densité de probabilité dnorm(t) correspondant aux valeurs **positives t** de la loi normale centrée réduite $T = N_{0,1}$ variant de dixième en dixième :

$t = 0.00 ; 0.01 ; \dots$. Les unités se lisent en ligne et les dixièmes en colonne.

Valeurs de t négatives : En raison de la symétrie de la courbe de densité, la table permet de déterminer les densités correspondant aux valeurs négatives de t :

$$dnorm(-t) = dnorm(t)$$

b) Table de la fonction de répartition pnorm : descriptif

Cette table donne pour toute valeur **positive t** de la variable normale centrée réduite

$T = N_{0,1}$ la valeur correspondante de la fonction de répartition pnorm(t, 0, 1) représentée par l'aire hachurée, qui est égale à la probabilité que T soit inférieur ou égal à t. (Annexe page 37)

$$pnorm(t, 0, 1) = p(T \leq t)$$

Les valeurs de t varient de centième en centième : $t = 0.00 ; 0.01 ; \dots$. Les unités et les dixièmes se lisent en ligne et les centièmes en colonne.

Exemple : Calculer la probabilité que $T = N_{0,1}$ soit inférieur ou égal à 1.32

$$pnorm(1.32, 0, 1) = \dots$$

Probabilité pour que $T = N_{0,1}$ soit supérieur à t, réel positif ou nul

L'aire totale comprise entre la courbe et l'axe des abscisses représente la somme des probabilités de la loi normale et est égale à 1. On a donc :

$$p(T > t) = 1 - p(T \leq t) = 1 - pnorm(t, 0, 1)$$

Autres probabilités

$$p(|T| \leq t) = 2 \operatorname{pnorm}(t, 0, 1) - 1$$

$$p(|T| > t) = 2(1 - \operatorname{pnorm}(t, 0, 1))$$

$$p(t_1 \leq T \leq t_2) = \operatorname{pnorm}(t_2, 0, 1) - \operatorname{pnorm}(t_1, 0, 1)$$

$$p(T \leq -t) = 1 - \operatorname{pnorm}(t, 0, 1)$$

II) La loi log-normale

1) Définition

Définition : Une variable aléatoire suit une loi **log-normale** si **son logarithme népérien suit une loi normale**.

Cette loi est très répandue, notamment dans le domaine des phénomènes socio-économiques. Chaque fois que les causes de variabilité, conformes aux conditions d'application du théorème Central-Limit sont multiplicatives, et non additives, le phénomène observé tend à suivre une loi log-normale.

Les distributions de salaires, de chiffres d'affaires suivent une loi log-normale.

Notation : On caractérise une variable aléatoire log-normale X par les paramètres μ et σ de la loi normale que suit son logarithme népérien.

$$\ln X = N_{\mu, \sigma} \quad \text{ou} \quad X = LN_{\mu, \sigma}$$

où μ et σ sont la moyenne et l'écart-type du logarithme népérien de X.

$T = \frac{\ln X - \mu}{\sigma}$ suit une loi $N_{0,1}$. On a : $X = e^{\sigma T + \mu}$

2) Caractéristiques de la loi log-normale

a) L'espérance mathématique

Définition : **L'espérance mathématique** de la loi $LN_{\mu, \sigma}$ est

$$E(LN_{\mu, \sigma}) = e^{\mu + \frac{\sigma^2}{2}}$$

b) La variance et l'écart-type

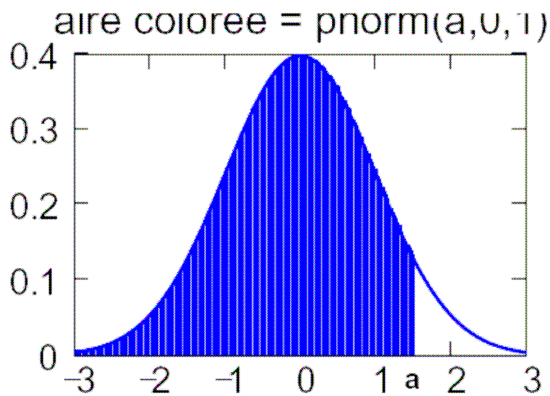
Définition : **La variance** de la loi $LN_{\mu, \sigma}$ est

$$V(LN_{\mu, \sigma}) = e^{2(\mu + \sigma^2)} (1 - e^{-\sigma^2})$$

L'écart-type de la loi $LN_{\mu, \sigma}$ est

$$\sigma(LN_{\mu, \sigma}) = \sqrt{e^{2(\mu + \sigma^2)} (1 - e^{-\sigma^2})}$$

Table extraite du cours de l'UPPA : Professeur : Marc Artzrouni



ANNEXE: Loi normale centrée-réduite: fonction de répartition pnorm

La table c-dessous donne l'aire colorée $\text{pnorm}(a,0,1)$ à l'intersection (col,li) où $a > 0$ s'écrit $a = li + col$ avec $li = \text{unités} + \text{dixièmes}$ et $col = \text{centième}$.

Exemple: si $a = 1.13$, $li = 1.1$, $col = 0.03$ et on lit $\text{pnorm}(1.13, 0, 1) = 0.8708$;

NB: par symétrie $\text{pnorm}(-a, 0, 1) = 1 - \text{pnorm}(a, 0, 1)$

M =

0	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	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	0.591	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	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	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	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	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	0.834	0.8365	0.8389
1	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	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	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	0.937	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	0.975	0.9756	0.9761	0.9767
2	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	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
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	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	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	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	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	0.998	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999

Fin