Ch. 4 : Variables aléatoires continues, loi normale.

1 Préliminaire : rappel sur le calcul intégral.

Pour généraliser les définitions du chapitre précédent à une variable aléatoire continue, il est nécessaire de sommer sur une infinité de valeurs x, c'est-à-dire de remplacer le signe $\sum_{i=a}^{i=b} f(i)$ par $\int_{x=a}^{x=b} f(x)dx$. Que signifie-t-il ?

Remarquons que si on place au dessus de chaque entier i un rectangle de largeur 1 et de hauteur $f(i)$, la somme $\sum_{i=a}^{b} f(i) \times 1$ est l'aire de la réunion de ces rectangles, donc l'aire délimitée par l'axe horizontal, le graphe d'une fonction en escalier et deux verticales. Plus généralement, si f est une fonction continue par morceaux, positive, sur $[a,b]$, le nombre $\int_{x=a}^{x=b} f(x)dx$ désigne l'aire de la région délimitée par le graphe de f l'horizontale $y=0$ et les deux verticales $x=a$ et $x=b$:

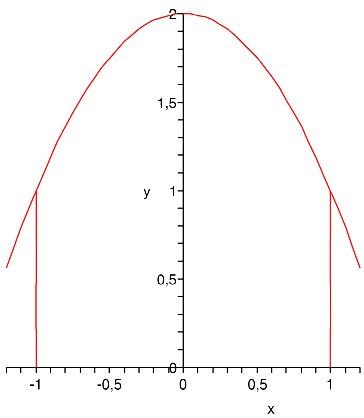

FIG. 1 – Aire correspondant à $\int_{x=-1}^{x=1}(2-x^2)dx$

Remarques :
- Cela a encore un sens si f prend des valeurs négatives, à condition de compter négativement les aires en-dessous de l'axe des abscisses.
- Tout ceci garde parfois un sens lorsque l'une ou les deux extrémités sont infinies : l'aire est obtenue en prenant la limite quand a tend vers $-\infty$ ou b vers $+\infty$. Il faut alors que cette limite existe et soit finie.

Comment calculer cette aire ? On considère une primitive F de f, c'est-à-dire une fonction dont la dérivée est f. Il en existe toujours si f est continue. Elle est unique à l'addition près d'une constante, et vraiment unique si on impose que $F(a)$ prenne une valeur donnée.

Exercice 1. *Soit $f : x \mapsto f(x) = x^2$.*
- Quelles sont les primitives de f ?
- Quelle est la primitive F de f qui vérifie $F(1) = 2$?

On a le théorème important suivant : pour toute primitive F de f,

$$\int_{x=a}^{x=b} f(x)dx = F(b) - F(a).$$

Il ramène le calcul d'une aire à un calcul de primitives.

Comment calculer une primitive ? (revoir votre cours de Terminale). Il faut d'abord connaitre les primitives des fonctions usuelles. Exemples :
- si $f(x) = x^n$, $(n \neq -1)$, $F(x) = \frac{x^{n+1}}{n+1} + c$ (où c est une constante quelconque).
- si $f(x) = 1/x$, $F(x) = \ln(x) + c$.
- si $f(x) = e^{kx}$, $F(x) = \frac{e^{kx}}{k} + c$.

Pour les fonctions plus compliquées, on s'y ramène par combinaison linéaire (facile), intégration par partie (pour les produits), ou par changement de variable (on pose $X = u(x)$, d'où $dX = u'(x)dx$. Si $f(x) = g(X)$, on a alors $\int_{x=a}^{x=b} f(x)dx = \int_{X=u(a)}^{X=u(b)} g(X)dX$). Néanmoins, dans certains cas, aucune de ces méthodes ne s'applique, et on ne peut trouver d'expression explicite de la primitive.

Exercice 2. *Déterminer l'aire délimitée par la parabole d'équation $y = 2 - x^2$, les deux verticales $x = -1$ et $x = 1$ et l'axe horizontal $y = 0$, représentée en figure 1.*

2 Densité et fonction de répartition d'une variable aléatoire continue.

La probabilité qu'une ampoule meure exactement à une heure très précise donnée est nulle. Plus généralement si X est une variable aléatoire continue, $p(X = a) = 0$. Pour une variable aléatoire continue, on ne s'intéresse qu'à la probabilité que le résultat x de X tombe dans un intervalle $[a, b]$: $p(a < X < b)$. On fait l'hypothèse suivante : il existe une fonction f telle que

$$p(a < X < b) = \int_{x=a}^{x=b} f(x)dx.$$

Cette fonction f est appelée **densité de probabilité**. Elle doit être positive (pour que p le soit), continue (pour admettre une primitive) et doit vérifier $\int_{x=-\infty}^{+\infty} f(x)dx = 1$ pour que $p(\Omega) = 1$.

Fonction de répartition de X. C'est l'application qui à tout c associe :

$$F(c) = p(X < c) = \int_{x=-\infty}^{c} f(x)dx.$$

Elle est croissante de 0 à 1, et permet de connaitre toutes les probabilités de tomber dans n'importe quel intervalle puisque :

$$p(X > a) = 1 - p(X < a) = 1 - F(a) \text{ et } p(a < X < b) = F(b) - F(a).$$

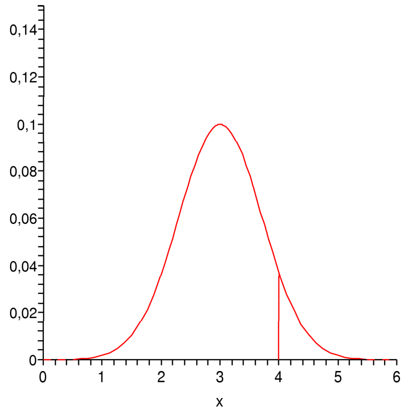

FIG. 2 – $p(X < 4)$, $p(X > 4)$.

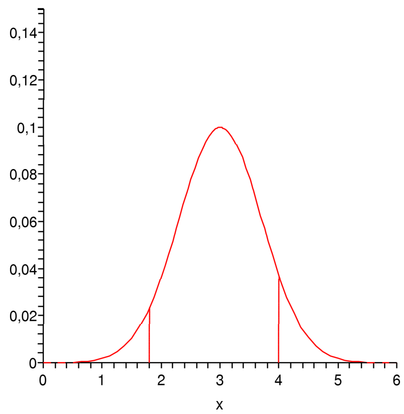

FIG. 3 – $p(1.8 < X < 4)$.

Remarque. Remplacer les inégalités strictes par des inégalités larges ne change rien puisque pour une variable aléatoire *continue*, la probabilité $p(X = a)$ que X prenne exactement la valeur a est nulle.

3 Paramètres d'une variable aléatoire continue.

Ils se définissent de manière analogue au cas discret en remplaçant la sommation finie par une intégration et en remplaçant p_i (probabilité que X prenne la valeur x_i) par $f(x)dx$ (probabilité que X prenne ses valeurs dans un tout petit intervalle dx) :

Espérance mathématique :

$$E(X) = \int_{-\infty}^{+\infty} x \ f(x)dx.$$

Variance et écart-type : Posons $\overline{x} = E(X)$. L'écart-type de X est donné par

$$\sigma(X) = \sqrt{\int_{-\infty}^{+\infty} (x - \overline{x})^2 \ f(x)dx}.$$

La variance est le carré de l'écart-type. On a encore (formule de Koenigs) :

$$\sigma(X)^2 = (\int_{-\infty}^{+\infty} x^2 \ f(x)dx) - \overline{x}^2.$$

4 Exemples de densités de probabilités usuelles

Loi uniforme. La probabilité que la valeur de X tombe dans un intervalle de longueur fixée reste la même quelquesoit cet intervalle inclus dans $[a, b]$. On doit alors avoir $f(x)$ constante, donc $f(x) = \frac{1}{b-a}$ sur $[a, b]$, et 0 en dehors, afin d'obtenir une aire totale de 1 :

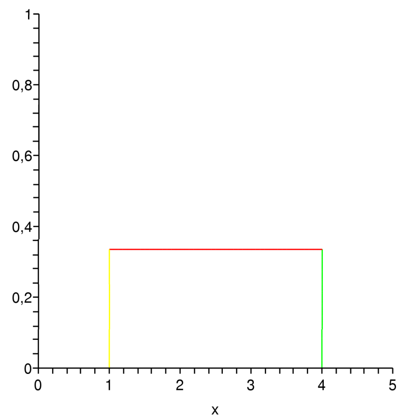

FIG. 4 – densités de la loi uniforme sur l'intervalle $[1, 4]$.

Exercice 3. *Pour tout c entre a et b, déterminer (par le graphique ou par le calcul) la probabilité : $p(X < c)$.*

On a $E(X) = \frac{a+b}{2}$ (pourquoi ?) et $\sigma(X)^2 = \frac{(b-a)^2}{12}$, ce qu'on peut vérifier par le calcul.

Loi exponentielle. Elle est souvent utilisée pour des questions de fiabilité (ex : durée de vie de circuits électroniques). Elle est définie pour $x \geq 0$ par $f(x) = \frac{1}{a}e^{-x/a}$, et 0 pour $x < 0$:

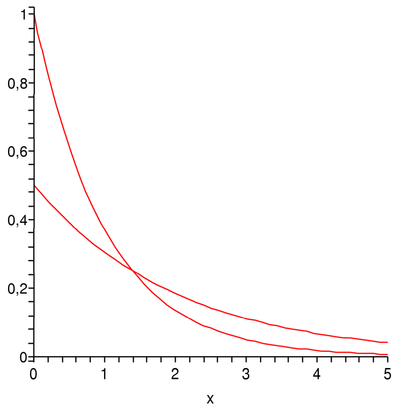

FIG. 5 – densités des lois exponentielles pour $a = 1$ ($f(0) = 1$) et $a = 2$ ($f(0) = 1/2$).

Exercice 4. *Si X suit la loi exponentielle de paramètre $a = 1$, déterminer $p(X < 2)$.*

On a $E(X) = a$ et $\sigma(X) = a$ (exercice de calcul intégral : démontrez-le). En particulier, si on connait l'espérance d'une loi exponentielle, on connait son paramètre a.

5 Loi normale (ou loi de Gauss)

Elle est souvent un bon modèle pour représenter une distribution de mesures physiques avec erreurs autour de la "valeur vraie" μ. Elle joue de plus un rôle central en statistiques pour des raisons que nous verrons dans les prochains chapitres : approximations, échantillonnages.

La densité de probabilité de cette loi a une forme de cloche symétrique autour d'une valeur moyenne μ :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

On a : $E(X) = \mu$ et $\sigma(X) = \sigma$. On la note $\mathcal{N}(\mu, \sigma)$. En particulier la loi normale centrée réduite $\mathcal{N}(0, 1)$ a pour densité :

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

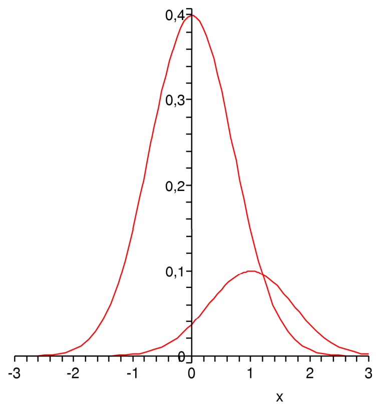

FIG. 6 – Graphes de la loi normale centrée réduite et de la loi normale $\mathcal{N}(1,4)$.

Remarque importante pour les calculs. Si Z suit une loi normale centrée réduite, on calcule sa loi de probabilité à l'aide de la table de sa fonction de répartition Π (voir la dernière page de ce chapitre). De plus, la courbe de densité la loi normale centrée réduite étant symétrique par rapport à l'axe vertical, on utilisera les propriétés

$$\begin{aligned} p(Z < -a) &= \Pi(-a) = 1 - \Pi(a) \\ p(|Z| < a) &= p(-a < Z < a) = \Pi(a) - \Pi(-a) = 2\Pi(a) - 1. \end{aligned}$$

Exercice 5. On suppose que Z suit la loi normale centrée réduite. A l'aide de la table de la fonction de répartition Π de Z (voir la dernière page de ce chapitre), déterminer :
- $p(Z < 2,28)$;
- $p(Z > 1,75)$;
- $p(1 < Z < 2)$;
- $p(Z < -1,28)$;
- $p(Z > -1,28)$;
- Trouver a tel que $p(|Z| < a) = 80\%$.

Si une variable aléatoire X suit $\mathcal{N}(\mu, \sigma)$, on peut calculer sa loi de probabilité à l'aide de la table de $\mathcal{N}(0,1)$ en centrant et réduisant, c'est-à-dire en posant :

$$p(X < a) = p(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}) = p(Z < A)$$

où $A = \frac{a-\mu}{\sigma}$ et où Z suit $\mathcal{N}(0,1)$.

Exercice 6. *On suppose que X suit la loi normale $\mathcal{N}(1,2)$.*
Déterminer la probabilité : $p(X < 1,65)$.

6 Construction de nouvelles variables aléatoires.

Etant donnée deux variables aléatoires (discrètes ou continues) à valeurs réelles, on peut (comme pour toute fonction !) construire leur somme $X_1 + X_2$, et leur produit X_1X_2. On peut aussi multiplier une variable aléatoire X par un scalaire : λX. On fait à chaque fois l'opération correspondante sur les valeurs.

Comme pour les distributions statistiques, l'espérance de la somme est la somme des espérances :

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

La variance de la somme de deux variables aléatoires discrètes est donnée par

$$\sigma^2(X_1 + X_2) = \sigma^2(X_1) + \sigma^2(X_2) + 2\text{cov}(X_1, X_2).$$

Donc si elles sont *indépendantes*, c'est la somme des variances. Ce résultat reste vrai pour les couples de variables aléatoires continues pour lesquelles on peut définir une covariance (cela nécessite de faire des intégrations sur deux variables).
On a aussi

$$E(\lambda X) = \lambda E(X) \text{ et } \sigma^2(\lambda X) = \lambda^2\sigma^2(X).$$

Connaissant les lois de probabilités de chacune des X_i il n'est pas évident de calculer la loi de probabilité de la variable somme ou produit. Cependant on connait cette loi dans les situations suivantes :

1. Si les n variables X_i suivent toutes la loi de Bernouilli (exemple : un jet de pièce, $X_i = 1$ si c'est "pile")) et sont indépendantes, la variable somme $X_1 + \cdots + X_n$ (n jets de pièces et on compte le nombre de "pile") suit la loi binomiale (voir le chapitre 2).
2. Si les n variables indépendantes X_i suivent toutes des lois normales $\mathcal{N}(\mu_i, \sigma_i)$, la variable somme $X_1 + \cdots + X_n$ suit encore une loi normale (nécessairement d'espérance la somme des espérances et de variance la somme des variances).
3. Si les n variables indépendantes X_i suivent toutes des lois de Poisson $\mathcal{P}(\lambda_i)$, la variable somme $X_1 + \cdots + X_n$ suit encore une loi de Poisson (nécessairement de paramètre $\sum \lambda_i$).
4. Si les n variables indépendantes X_i suivent toutes la loi normale centrée réduite $\mathcal{N}(0, 1)$, la variable $Y = X_1^2 + \cdots X_n^2$ suit une loi appelée "**loi du Chi-carré à n degré de liberté**" (loi du $\chi^2(n)$.) Ce n'est plus une loi normale. Cette loi est calculée à l'aide de tables. Elle est très utile en théorie des tests statistiques.
5. Si X suit la loi normale centrée réduite et Y la loi du $\chi^2(n)$, $T := \frac{X}{\sqrt{Y/n}}$ suit une loi appelée **loi de Student à n degrés de liberté** (loi $\mathcal{S}(n)$). Elle est aussi tabulée et est utile en statistique (pour les petits échantillons).

Test sur le chapitre 4

1. Quelles conditions doit vérifier une fonction f pour être densité de probabilité d'une variable aléatoire continue ?
2. Quel est le lien entre densité et fonction de répartition ?
3. Connaissant la fonction de répartition F de X, comment calculer $p(a < X < b)$?
4. Quelles formules permettent de calculer l'espérance et la variance d'une variable aléatoire continue X de densité f ?
5. Quelle est la fonction densité de la loi normale centrée réduite ?
6. Si Z désigne une variable aléatoire suivant la loi normale centrée réduite, calculer $p(Z > 1)$.
7. A-t-on toujours la propriété suivante : "la variance de la somme de deux variables aléatoires et la somme des variances" ?

Chapitre 4 : Travaux dirigés

1. La durée de vie d'une ampoule électrique suit une loi exponentielle d'espérance 1000 heures. Déterminer la probabilité que la durée de vie d'une ampoule dépasse 2000 heures.

2. On suppose qu'une variable aléatoire X suit la loi normale $\mathcal{N}(m, \sigma)$. Déterminer (en se ramenant à la loi normale centrée réduite Z) :
   - $p(m - \sigma < X < m + \sigma)$;
   - $p(m - 2\sigma < X < m + 2\sigma)$.

3. *Modélisation par une loi normale.*
   L'éclairage d'une commune est assuré par 2000 lampadaires dont les lampes ont une durée de vie moyenne de 1000 heures. Les tests montrent que cette durée de vie suit une loi normale autour de cette valeur avec un écart-type de 200 heures. On demande de fournir aux services d'entretien une estimation probabiliste des données suivantes :
   (a) le nombre de lampes hors d'usage au bout de 700 heures ;
   (b) le nombre de lampes à remplacer entre la 900^{eme} et la 1000^{eme} heure ;
   (c) le nombre d'heures écoulées pour que 10% des lampes soit hors d'usage.

4. Une usine produit des bobines de fil pour l'industrie textile. Soit X la variable aléatoire qui à toute bobine extraite de la production associe sa longueur de fil en mètres. On admet que X soit la loi normale d'espérance 50 et d'écart-type 0,2. Calculer la probabilité de chacun des évènements suivants :
   (a) la longueur du fil de la bobine est inférieure à 50,19 m.
   (b) la longueur du fil de la bobine est supérieure à 50,16 m.
   (c) la longueur du fil de la bobine est comprise entre 50,16 m et 50,19 m.
   (d) Déterminer le nombre réel a tel que : $p(50 - a < X < 50 + a) = 90\%$.

Travail personnel :

1. On suppose qu'une variable aléatoire X suit la loi uniforme sur l'intervalle $[a, b]$. Vérifier que :

$$E(X) = \frac{a+b}{2}, \quad \sigma(X) = \sqrt{\frac{(b-a)^2}{12}}.$$

2. On suppose que X suit la loi exponentielle de paramètre a. Vérifier que :

$$E(X) = a, \quad \sigma(X) = a.$$

———————————————-

Intégrale $\Pi(t)$ de la Loi Normale Centrée Réduite $\mathcal{N}(0;\,1)$.

$$\Pi(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \quad \text{et} \quad \Pi(-t) = 1 - \Pi(t).$$

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
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000